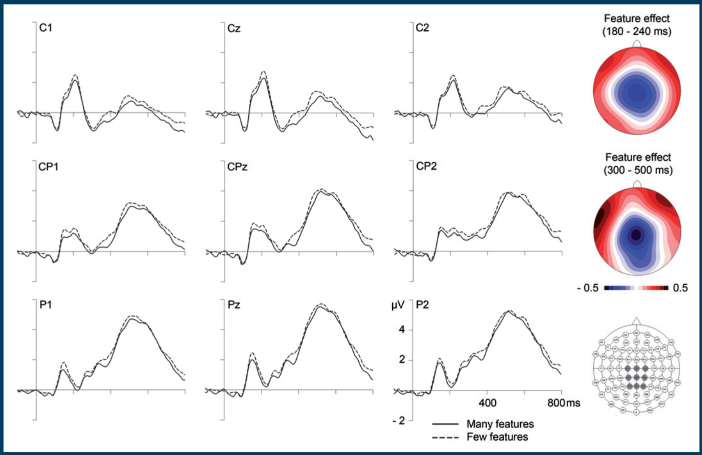
frontiers
RESEARCH TOPICS



MEANING IN MIND: SEMANTIC RICHNESS EFFECTS IN LANGUAGE PROCESSING

Topic Editors
Penny M. Pexman,
Paul D. Siakaluk and Melvin J. Yap

frontiers in
HUMAN NEUROSCIENCE

frontiers

ISSN 1664-8714
ISBN 978-2-88919-182-6
DOI 10.3389/978-2-88919-182-6

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

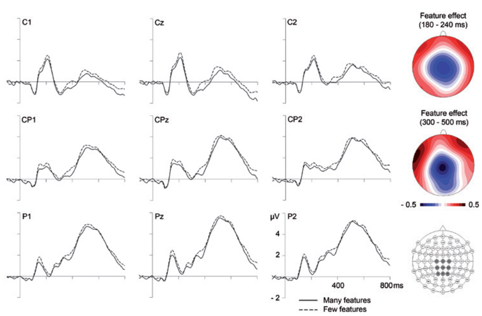
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

MEANING IN MIND: SEMANTIC RICHNESS EFFECTS IN LANGUAGE PROCESSING

Topic Editors:
Penny M. Pexman, Department of Psychology, University of Calgary
Paul D. Siakaluk, Department of Psychology, University of Northern British Columbia
Melvin J. Yap, Department of Psychology, National University of Singapore



Influences of the number of semantic features on event-related brain potentials at centro-parietal electrode sites.
And then smaller: On the right are topographical distributions of feature effects (many minus few semantic features) between 180 and 240 (top) and between 300 and 500 ms (middle), as well as a map of electrode locations with the depicted sites highlighted in dark gray.

To the skilled reader, the process by which meaning is extracted from print feels almost effortless. This feeling, however, is misleading, because the process is actually quite complex. As skilled readers we have little conscious insight about this process, so in order to study it researchers must devise careful experiments, and make use of all the tools available to modern Cognitive Science. The contributions in this book represent one such approach: the study of semantic richness effects provides important clues about how readers generate meaning from words.

Semantic richness refers to natural variability that exists in the meaning information associated with different words, and semantic richness effects are the behavioral and neural consequences of that variability. The contributions in this book illustrate that semantic richness can be defined in many different ways, as a function of every extant model of word meaning, and that semantic processing can be examined with varied behavioral, neuroimaging, and neuropsychological paradigms. As such, these contributions test current models and provide new answers to the complex question of how word meaning is understood.

Table of Contents

frontiers in
HUMAN NEUROSCIENCE

EDITORIAL
published: 05 November 2013
doi: 10.3389/fnhum.2013.00723

Introduction to the research topic meaning in mind: semantic richness effects in language processing

Penny M. Pexman[1]*, Paul D. Siakaluk[2] and Melvin J. Yap[3]

[1] *Department of Psychology, University of Calgary, Calgary, AB, Canada*
[2] *Department of Psychology, University of Northern British Columbia, Prince George, BC, Canada*
[3] *Department of Psychology, National University of Singapore, Singapore, Singapore*
**Correspondence: pexman@ucalgary.ca*

Edited by:
Hauke R. Heekeren, Freie Universität Berlin, Germany

Keywords: semantic processing, word meaning, semantic richness, embodied cognition, concrete concepts, abstract concepts, lexical processing

The ultimate goal of reading is to extract meaning from printed words. However, the mechanisms that mediate orthography and semantics are not well-understood, and have rarely been implemented in computational models. To address this puzzle, one of the strategies cognitive scientists have begun to use is to examine semantic richness effects. Semantic richness effects refer to the finding that words associated with relatively more semantic information are recognized faster and more accurately, due to their possessing richer, better-specified semantic representations. Importantly, semantic richness is not a unitary concept. Instead, it draws on various theoretical perspectives and can vary along multiple dimensions. Thus, by examining which dimensions of semantic richness influence visual word recognition behavior, we gain insight about which theoretical perspectives seem to be promising descriptions of the process by which meaning is extracted from print. Our goal for this Frontiers Research Topic was to highlight the latest findings regarding semantic richness and theoretical developments on the issue of semantic processing. Our hope was to provide a forum for state-of-the-art research in this field, and to foster new theoretical advances. The 17 contributions that comprise the Research Topic certainly represent the state of the art; methodologies include ERP, fMRI, TMS, and behavioral approaches, and involve both intact and patient populations. Together, these contributions give rise to a number of inferences about semantic richness effects and implications of those effects for our understanding of semantic processing effects in visual word recognition.

MEANING IS MULTIDIMENSIONAL

The Research Topic contributions build on previous literature, providing further empirical support for several semantic richness dimensions and the frameworks from which those dimensions are derived. Gould et al. (2012); Recchia and Jones (2012); Yap et al. (2012) report semantic neighborhood effects (faster responses for words with more semantic neighbors or denser semantic neighborhoods) in naming and lexical decision tasks, providing evidence that lexical co-occurrence is an important dimension in semantic memory. Hargreaves and Pexman (2012); Taler et al. (2013) show that lexical decision performance is facilitated for words with more meaning senses, providing support for the notion that meaning information is represented in a distributed fashion. The typicality effects reported by Woollams (2012) support the claim that words' feature structure is important to semantic memory. Further, Recchia and Jones (2012); Yap et al. (2012) show that words that generate more features in feature listing tasks produce faster naming, lexical decision, and semantic categorization responses, Hargreaves et al. (2012a) report that those words are also better remembered in free recall. Finally, there is evidence supporting embodied frameworks of semantic memory from studies reported by Esopenko et al. (2012); McNorgan (2012). Further support for the embodied framework is provided by Hansen et al. (2012); Hargreaves et al. (2012b); Newcombe et al. (2012); Tousignant and Pexman (2012); Yap et al. (2012), as all of these studies report body-object interaction effects (faster processing for words that refer to objects the human body can easily interact with) in tasks that include naming, lexical decision, and semantic categorization. Convergent evidence that perceptual and sensorimotor information are important dimensions of meaning comes from the observations of Hargreaves and Pexman (2012); Newcombe et al. (2012); Yap et al. (2012) by which imageability effects (faster responses for words that are associated with imagery) are reported in a number of word recognition tasks.

In addition, in the contributions of Hargreaves and Pexman (2012); Newcombe et al. (2012); Recchia and Jones (2012); Yap et al. (2012) there are demonstrations that multiple semantic richness effects can be observed simultaneously, suggesting that each richness dimension explains unique variance in word recognition behavior. The implication is that no single dimension (and associated framework) will be sufficient to explain the process by which meaning is derived from print. Instead, as argued by Dilkina and Lambon Ralph (2013); Jones and Golonka (2012); Kalénine et al. (2012), semantic memory is multidimensional.

SEMANTIC PROCESSING IS VARIABLE AND DYNAMIC

The findings of Kalénine et al. (2012); Woollams (2012) support the inference that semantic processing is variable as a function of disease. By studying the dimensions of meaning that are more resistant to brain damage these studies provide important new clues about the structure of meaning in the mind. The contributions of Hargreaves and Pexman (2012); Hansen et al. (2012) show that semantic processing is variable as a function of both short-term and long-term experience. Further variability is revealed in Jones and Golonka (2012); Kalénine et al. (2012); Rabovsky et al. (2012); Taler et al. (2013), where the timecourse of processing is examined in order to dissociate richness dimensions. Results show, first, that semantic information is generated quite early in the process of word recognition and, second, that different dimensions of meaning may be influential at different times as semantic processing unfolds.

Contributions by Gould et al. (2012); Hansen et al. (2012); Hargreaves and Pexman (2012); Recchia and Jones (2012); Tousignant and Pexman (2012); Yap et al. (2012) demonstrate that the process of generating meaning from print is a dynamic one, where contextual factors like task demands shape the information that is generated from letter strings. These demonstrations are consistent with the notion of a flexible lexical processor (Balota and Yap, 2006) that is sensitive to task contexts so as to optimize task performance via attentional control. The present findings also permit the inference that the semantic richness effects observed in a given task do not provide veridical insight about static semantic representations. Semantic representation is not fixed and so cannot be revealed in a single task or context (Kiefer and Pulvermüller, 2012). Rather, meaning is actively constructed and shaped to meet task demands. Dimensions that are important in one context may not be important in others. Certainly, it now seems clear that there are many candidate dimensions of meaning, but the context will dictate the actual effects observed.

FUTURE DIRECTIONS: ABSTRACT MEANING AND OTHER CHALLENGES

As has been typical in the lexical semantic literature, most of the contributions in this Research Topic focus on semantic processing of concrete words, like TRUCK, where the word refers to an object or entity in the world. As such, while we know quite a lot about how concrete meanings might be processed, we know much less about how abstract meanings are understood. This is problematic because abstract words make up a large part of the average person's vocabulary; the focus on concrete word meaning creates a situation where we are studying only part of the human lexicon. In two of the present papers, however, the authors use semantic richness effects to begin to study semantic processing of abstract words, like TRUTH. Newcombe et al. (2012); Recchia and Jones (2012) explore semantic richness dimensions that could be relevant to abstract word meaning. Since many of the richness dimensions that are influential for concrete words are not as relevant to the meanings of abstract words (e.g., those dimensions that refer to objects), the richness dimensions that influence abstract word meaning are somewhat different. For instance, Newcombe et al. (2012) show that while body-object interaction is an important dimension for concrete words, emotion information is important for abstract words, consistent with predictions derived from the embodied cognition framework of Kousta et al. (2010). In addition, Recchia and Jones (2012) show that richer linguistic contexts (larger semantic neighborhoods) facilitate abstract word processing. These contributions are first steps in the study of abstract word meaning, and this issue will need to be taken up in future research.

We suggest, further, that future research on this topic should continue to explore several of the other important avenues opened here, for instance, the role of individual differences in semantic processing and the joint effects of different semantic richness dimensions. There are additional issues that have not yet received much attention but will be important; for instance, the issue of whether semantic richness dimensions influence processing in a linear or non-linear manner, and the extent to which richness effects extend beyond single-word contexts to influence processing of phrases and sentences. These and other research questions should be addressed in order that we are able to further refine our understanding of how word meaning is processed in mind and brain.

CONCLUSION

The contributions in this Frontiers Research Topic highlight a number of dimensions of semantic richness and the contexts in which they are observed. The contributions cohere around several insights: multiple types of information are constitutive of word meaning, and semantic processing is a dynamic process that must be tracked with careful consideration of context and other sources of variability; the challenges for theories of semantic meaning are to capture this multidimensionality, and to extend their reach to include abstract meanings.

REFERENCES

Balota, D. A., and Yap, M. J. (2006). "Attentional control and flexible lexical processing: explorations of the magic moment of word recognition," in *From Inkmarks to Ideas: Current Issues in Lexical Processing*, ed S. Andrews (New York, NY: Psychology Press), 229–258.

Dilkina, K., and Lambon Ralph, M. A. (2013). Conceptual structure within and between modalities. *Front. Hum. Neurosci.* 6:333. doi: 10.3389/fnhum.2012.00333

Esopenko, C., Gould, L., Cummine, J., Sarty, G. E., Kuhlmann, N., and Borowsky, R. (2012). A neuroanatomical examination of embodied cognition: semantic generation to action-related stimuli. *Front. Hum. Neurosci.* 6:84. doi: 10.3389/fnhum.2012.00084

Gould, L., Cummine, J., and Borowsky, R. (2012). The cognitive chronometric architecture of reading aloud: semantic and lexical effects on naming onset and duration. *Front. Hum. Neurosci.* 6:287. doi: 10.3389/fnhum.2012.00287

Hargreaves, I. S., and Pexman, P. M. (2012). Does richness lose its luster? Effects of extensive practice on semantic richness in visual word recognition. *Front. Hum. Neurosci.* 6:234. doi: 10.3389/fnhum.2012.00234

Hargreaves, I. S., Pexman, P. M., Johnson, J. C., and Zdrazilova, L. (2012a). Richer concepts are better remembered: number of features effects in free recall. *Front. Hum. Neurosci.* 6:73. doi: 10.3389/fnhum.2012.00073

Hargreaves, I. S., Leonard, G. A., Pexman, P. M., Pittman, D. J., Siakaluk, P. D., and Goodyear, B. G. (2012b). The neural correlates of the body-object interaction effect in semantic processing. *Front. Hum. Neurosci.* 6:22. doi: 10.3389/fnhum.2012.00022

Hansen, D., Siakaluk, P. D., and Pexman, P. M. (2012). The influence of print exposure on the body-object interaction effect in visual word recognition. *Front. Hum. Neurosci.* 6:113. doi: 10.3389/fnhum.2012.00113

Jones, L. L., and Golonka, S. (2012). Different influences on lexical priming for integrative, thematic, and taxonomic relations. *Front. Hum. Neurosci.* 6:205. doi: 10.3389/fnhum.2012.00205

Kalénine, S., Mirman, D., and Buxbaum, L. J. (2012). A combination of thematic and similarity-based semantic processes confers resistance to deficit following left hemisphere stroke. *Front. Hum. Neurosci.* 6:106. doi: 10.3389/fnhum.2012.00106

Kiefer, M., and Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex* 48, 805–825. doi: 10.1016/j.cortex.2011.04.006

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., and Del Campo, E. (2010). The representation of abstract words: why emotion matters. *J. Exp. Psychol. Gen.* 140, 14–34. doi: 10.1037/a0021446

McNorgan, C. (2012). A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Front. Hum. Neurosci.* 6:285. doi: 10.3389/fnhum.2012.00285

Newcombe, P. I., Campbell, C., Siakaluk, P. D., and Pexman, P. M. (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. *Front. Hum. Neurosci.* 6:275. doi: 10.3389/fnhum.2012.00275

Rabovsky, M., Sommer, W., and Abdel Rahman, R. (2012). The time course of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:11. doi: 10.3389/fnhum.2012.00011

Recchia, G., and Jones, M. N. (2012). The semantic richness of abstract concepts. *Front. Hum. Neurosci.* 6:315. doi: 10.3389/fnhum.2012.00315

Taler, V., Kousaie, S., and López Zunini, R. (2013). ERP measures of semantic richness: the case of multiple senses. *Front. Hum. Neurosci.* 7:5. doi: 10.3389/fnhum.2013.00005

Tousignant, C., and Pexman, P. M. (2012). Flexible recruitment of semantic richness: context modulates body-object interaction effects in lexical-semantic processing. *Front. Hum. Neurosci.* 6:53. doi: 10.3389/fnhum.2012.00053

Woollams, A. M. (2012). Apples are not the only fruit: the effects of concept typicality on semantic representation in the anterior temporal lobe. *Front. Hum. Neurosci.* 6:85. doi: 10.3389/fnhum.2012.00085

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., and Huff, M. J. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:72. doi: 10.3389/fnhum.2012.00072

Received: 18 September 2013; accepted: 11 October 2013; published online: 05 November 2013.

*Citation: Pexman PM, Siakaluk PD and Yap MJ (2013) Introduction to the research topic meaning in mind: semantic richness effects in language processing. Front. Hum. Neurosci. **7**:723. doi: 10.3389/fnhum.2013.00723*

This article was submitted to the journal Frontiers in Human Neuroscience.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 17 April 2012
doi: 10.3389/fnhum.2012.00072

An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition

Melvin J. Yap[1]*, Penny M. Pexman[2], Michele Wellsby[2], Ian S. Hargreaves[2] and Mark J. Huff[2]

[1] Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Singapore
[2] Department of Psychology, University of Calgary, Calgary, AB, Canada

Edited by:
Paul D. Siakaluk, University of Northern British Columbia, Canada

Reviewed by:
Jon Andoni Duñabeitia, Basque Center on Cognition, Brain and Language, Spain
David Balota, Washington University, USA
Michael J. Cortese, University of Nebraska at Omaha, USA

****Correspondence:***
Melvin J. Yap, Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Block AS4, #02-07, Singapore 117570.
e-mail: melvin@nus.edu.sg

There is considerable evidence (e.g., Pexman et al., 2008) that semantically rich words, which are associated with relatively more semantic information, are recognized faster across different lexical processing tasks. The present study extends this earlier work by providing the most comprehensive evaluation to date of semantic richness effects on visual word recognition performance. Specifically, using mixed effects analyses to control for the influence of correlated lexical variables, we considered the impact of number of features, number of senses, semantic neighborhood density, imageability, and body–object interaction across five visual word recognition tasks: standard lexical decision, go/no-go lexical decision, speeded pronunciation, progressive demasking, and semantic classification. Semantic richness effects could be reliably detected in all tasks of lexical processing, indicating that semantic representations, particularly their imaginal and featural aspects, play a fundamental role in visual word recognition. However, there was also evidence that the strength of certain richness effects could be flexibly and adaptively modulated by task demands, consistent with an intriguing interplay between task-specific mechanisms and differentiated semantic processing.

Keywords: semantic richness, visual word recognition, imageability, semantic neighborhood density, body-object interaction, semantic classification, lexical decision, progressive demasking

Although the ultimate goal of reading is to extract meaning from visually printed words, the effect of meaning-level influences on lexical processing is surprisingly poorly understood (see Pexman, in press, for a recent review; see also Balota et al., 1991). For the most part, the empirical literature has focused on how sublexical (see Carreiras and Grainger, 2004) and lexical (see Balota et al., 2006) representations influence visual word recognition. Likewise, despite their complexity and theoretical sophistication, influential computational models of visual word recognition (e.g., Perry et al., 2007) are silent on the role of semantic information (but see Harm and Seidenberg, 2004). Indeed, semantic effects are difficult to reconcile with classic logogen-based models of word recognition, which implicitly assume that lexical processing latencies tap a *magic moment* (Balota, 1990), i.e., a discrete moment in time when the lexical entry for a word has been identified but meaning-level information has not yet been accessed.

A number of studies (e.g., Buchanan et al., 2001; Cortese and Fugett, 2004; Duñabeitia et al., 2008; Pexman et al., 2008; Siakaluk et al., 2008; Yap et al., 2011) suggests that a word associated with relatively more semantic information can be considered to be semantically *richer*, and word recognition is generally facilitated for such words. A number of dimensions have been identified that appear to tap a word's semantic richness, including: (1) the *number of features* (NF) associated with its referent, (2) its *semantic neighborhood density* (SND) in high-dimensional semantic space, (3) the *number of distinct first associates* (NoA) elicited by the word in a free-association task (Nelson et al., 1998), (4) *imageability*, the extent to which a word evokes mental imagery of things or events (imageability), (5) *number of senses* (NS), the number of meanings associated with a word, and (6) *body–object interaction* (BOI), the extent to which a human body can physically interact with the word's referent. Specifically, word recognition is typically faster for words when their referents are associated with many semantic features (Pexman et al., 2003, 2008), when they are located in dense semantic neighborhoods (Buchanan et al., 2001; Shaoul and Westbury, 2010), when they elicit many associates (Duñabeitia et al., 2008), when they evoke more imagery (Cortese and Fugett, 2004), when they possess multiple meanings (Yap et al., 2011), and when they evoke more sensorimotor information (Siakaluk et al., 2008). That each of these variables affects word recognition behavior suggests that each taps an important aspect of semantic representation.

These findings can be accommodated by the embellished interactive activation framework suggested by Balota (1990; see also Balota et al., 1991), where there is bidirectional cascaded processing between letter-level, lexical-level, and semantic-level representations. Importantly, one could assume that there is stronger top-down feedback from semantic-level to orthographic-level representations for semantically rich words, facilitating lexical access for such words. The interactive activation framework is based on the assumption that lexical processing is subserved by mental lexicons containing localist units that represent the spelling, sound, and meaning of a word (see Coltheart et al., 2001), and that orthographic and semantic processing are handled by separate systems.

An important theoretical alternative to localist models is represented by parallel distributed processing (PDP) models, which do not assume the existence of mental lexicons. Instead, orthographic, phonological, and semantic information are respectively represented by distributed patterns of activity over separate layers of orthographic, phonological, and semantic neuron-like processing units (Plaut et al., 1996). In the classic connectionist triangle model of lexical processing (e.g., Seidenberg and McClelland, 1989), when a word is presented, activation flows from the orthographic to the semantic layer (either directly or mediated by the phonological layer) via weighted connections that reflect the network's knowledge of the mappings between orthography and semantics. Semantically rich words, which possess more stable and more readily computable meanings (Strain et al., 1995), are represented by the activation of more semantic units, hence yielding greater feedback activation from the semantic to the orthographic layer; this facilitates word recognition (see Hino et al., 2002, for more discussion of how PDP models can handle semantic richness effects).

Importantly, within the PDP approach, there is no sharp delineation between lexical and semantic processing. Instead, word recognition is assumed to rely on a common cognitive system where lexical and semantic knowledge interdependently and concurrently influence word recognition (Dilkina et al., 2008, 2010). Indeed, this view meshes well with the literature on semantic dementia, a variant of frontotemporal lobe dementia marked by deficits in conceptual and lexical knowledge (see Hodges et al., 1998, for a review). Although the finer details of this literature are beyond the scope of the present report, there is broad support for a positive correlation between lexical and conceptual deficits in semantic dementia patients, which is consistent with a single system that mediates both lexical and semantic processing (see Dilkina et al., 2010, for a connectionist model of lexical/semantic processing that explains the patient data).

SEMANTIC RICHNESS EFFECTS ARE MULTIDIMENSIONAL AND TASK-DEPENDENT

Importantly, although all the measures described ostensibly reflect aspects of a word's semantic richness, it is clear that they do *not* tap a common undifferentiated construct. For example, NF and NS seem to primarily reflect the complexity of a word's semantic representation, while SND and NoA may tap the extent to which that representation is interconnected with those of other words. Likewise, while imageability effects (Strain et al., 1995; Cortese et al., 1997; Cortese and Fugett, 2004) could be mediated by the interactions between lexical and visual (Paivio, 1991) or contextual (Schwanenflugel, 1991) information, BOI effects appear to implicate embodied sensorimotor representations (Siakaluk et al., 2008; Tillotson et al., 2008; Bennett et al., 2011; Wellsby et al., 2011). Consistent with this, the bivariate correlations between the different semantic richness variables are relatively modest, and the measures are also able to account for unique variance in word recognition performance (Pexman et al., 2008).

More relevantly for the purposes of the present study, the effects of semantic richness dimensions are flexibly and adaptively modulated by the specific demands of different lexical processing tasks (see Balota and Yap, 2006). For example, semantic richness variables are better predictors of semantic classification, compared to lexical decision, performance. In semantic classification, participants are required to discriminate between words from different semantic categories (e.g., is CABBAGE concrete or abstract?), while in lexical decision, they have to discriminate between real words and made-up words (e.g., FLIRP). Semantic processing is implicated to a greater extent in semantic classification because participants have to resolve the *specific* meaning of a word in order to make a correct response, whereas they can rely heavily on familiarity-based information to drive a lexical decision response (Balota and Chumbley, 1984).

Similarly, semantic effects are typically stronger in lexical decision than in speeded pronunciation performance. For example, early work by Chumbley and Balota (1984) reported that semantic variables such as *instance dominance* (the likelihood that a word will be given as an example to a category in response to the category name), number of associates, and NS produced reliable effects in lexical decision, but not in speeded pronunciation. More recent studies employing larger sets of stimuli (e.g., Balota et al., 2004) indicate that semantic effects (e.g., imageability effects) *can* be reliably detected in the pronunciation task, although they are greatly attenuated (see also Yap and Balota, 2009). These findings suggest that semantic information plays a stronger role in lexical decision, compared to pronunciation, because semantic information can be recruited to drive the familiarity-based word/non-word discrimination process that is specific to lexical decision (Balota and Chumbley, 1984; Chumbley and Balota, 1984).

The multidimensionality and task-specificity of semantic richness effects are also evident at a more fine-grained level. For example, although high-NF words are recognized faster in both lexical decision and semantic classification (Pexman et al., 2008), a denser semantic neighborhood is associated with faster lexical decision performance, but has no effect on semantic classification performance (Yap et al., 2011). While lexical decision is facilitated by stronger semantics → orthography feedback for words from dense neighborhoods, the opposing effects of nearby (facilitatory) and distant (inhibitory) neighbors may cancel each other out (Mirman and Magnuson, 2006) in tasks which emphasize semantic processing (e.g., semantic classification).

Intriguingly, the effect of semantic ambiguity is also very different in lexical decision and semantic classification. Specifically, although there is an ambiguity *advantage* (i.e., better performance for words with many senses) in lexical decision (e.g., Borowsky and Masson, 1996; Hargreaves et al., 2011), there is either a null ambiguity effect or an ambiguity *disadvantage* in semantic classification (Piercey and Joordens, 2000; Hino et al., 2002, 2006; Pexman et al., 2004; Hargreaves et al., 2011; Yap et al., 2011). Similar to neighborhood density, multiple meanings produce greater semantic feedback, which is helpful for lexical decision. However, a task which requires more focus on semantic processing can be slowed down by ambiguity, due to one-to-many mappings between orthography and semantics for ambiguous words (Borowsky and Masson, 1996), by increased competition between the different activated meanings (Grainger et al., 2001), or by competition between the activated meanings and the required response (Pexman et al., 2004).

The foregoing findings underscore the importance of examining semantic richness effects across a constellation of lexical processing paradigms, as effects are not process-pure but instead reflect both task-specific and task-general processing (Balota and Chumbley, 1984; Grainger and Jacobs, 1996). In line with this principle, Pexman et al. (2008) used hierarchical multiple regression to explore the effects of various semantic richness dimensions on lexical decision and semantic classification performance, basing their analyses on a common set of 514 concrete words from the McRae et al. (2005) norms. Yap et al. (2011) extended this work by also considering richness effects on speeded pronunciation (i.e., read words aloud) performance. Collectively, these studies yielded a number of the intriguing task-specific findings discussed earlier. In the same vein, Duñabeitia et al. (2008) studied the effect of NoA across different visual word recognition tasks, including speeded pronunciation, lexical decision, eye movements during sentence reading (see Rayner, 1998, for a review), and progressive demasking (Dufau et al., 2008). Eye movement data provide on-line, moment-to-moment measures of cognitive processes implicated in reading, while progressive demasking is a relatively novel perceptual identification task where a word gradually emerges from a mask over time, and the time taken to identify the specific word being presented is measured.

THE PRESENT RESEARCH

The primary objective of the present study is to provide the most comprehensive evaluation to date of the impact of extant semantic richness dimensions (NF, SND, imageability, NS, BOI) across different visual word recognition tasks. In addition to the ubiquitous traditional tasks (i.e., lexical decision, speeded pronunciation, semantic classification) examined by Pexman et al. (2008) and Yap et al. (2011), we also include newer tasks such as the progressive demasking task (PDT; Dufau et al., 2008) and go/no-go lexical decision (Gordon, 1983; Perea et al., 2002). In the progressive masking task, a word stimulus (e.g., DOG) is rapidly alternated with a mask (e.g., ###), and through successive display changes, the word gradually emerges from the mask. Participants make a button press as soon as they can identify the stimulus, hence yielding response time (RT) measures for a perceptual identification paradigm based on the presentation of visually degraded stimuli. Since it is a perceptual identification task, one might argue that unique stimulus identification is mandatory in progressive demasking; hence, progressive demasking latencies might provide a purer measure of lexical processing (Carreiras et al., 1997). Methodologically speaking, progressive demasking has important advantages over lexical decision and speeded pronunciation, as the experimenter does not need to create non-word distracters and performance is unaffected by articulatory factors (Ferrand et al., 2011). Finally, because of the way it is set up, progressive demasking slows down and stretches out the recognition process, potentially making this task *more* sensitive to underlying perceptual (Dufau et al., 2008) and semantic (Ferrand et al., 2011) processing.

The go/no-go lexical decision task[1] is an interesting variation on the standard task in which participants respond by pressing a button when a word is presented but *withhold* their response when a non-word is presented. The go/no-go task possesses a number of advantages. According to Perea et al. (2002), in addition to yielding faster, more accurate, and less noisy performance, go/no-go lexical decision also reduces task-specific processing demands (e.g., response competition during the decision process). In order to rigorously rule out the influence of correlated variables, which may spuriously inflate the predictive power of semantic richness variables (see Gernsbacher, 1984), linear mixed effects analyses (Baayen et al., 2008) will be used in the present study to control for phonological onsets (Balota et al., 2004) and established lexical variables (Balota et al., 2004; Yap and Balota, 2009); linear mixed models also allow us to generalize across both participants and items using a single model. All five datasets (LDT, go/no-go LDT, pronunciation Task, PDT, and SCT) are new, collected for the purposes of the present paper, and details of those datasets will be provided in the Methods section.

METHOD

PARTICIPANTS

Participants in all five tasks were undergraduate students at the University of Calgary who received course credit for participating. All participants reported that English was their first language and that they had normal or corrected-to-normal vision. There were 38 participants in the SCT, 31 participants in the LDT, and 30 in each of the other tasks.

MATERIALS

The word stimuli for this study were 514 concrete words[2] from McRae et al.'s (2005) norms. The stimuli also included 514 non-words used in the LDT and go/no-go LDT (which were matched to the words for length), and 514 abstract words which served as fillers in the SCT. The variables in the analyses were divided into three clusters: surface, lexical, and semantic variables (see **Table 1** for descriptive statistics of predictors and measures for items included in the analyses).

Surface variables

Dichotomous variables were used to code the initial phoneme of each word (1 = presence of feature; 0 = absence of feature) on 13 features: affricative, alveolar, bilabial, dental, fricative, glottal, labiodental, liquid, nasal, palatal, stop, velar, and voiced (Balota et al., 2004). These control for the variance associated with voice key biases in speeded pronunciation.

Lexical variables

These included log frequency (Brysbaert and New, 2009), number of morphemes, and number of letters. In order to address the high correlations between orthographic (Coltheart et al., 1977) and phonological (Yates, 2005) neighborhood size ($r = 0.79$), and

[1]The go/no-go lexical decision task should be distinguished from the go/no-go speeded pronunciation task (e.g., Hino and Lupker, 1998, 2000). In the latter task, participants name a stimulus aloud only if it is a word and withhold their response if it is a non-word. One of the reviewers made the interesting suggestion that this task could potentially provide important new insights into the task-specificity of semantic effects. Specifically, in go/no-go pronunciation, the response is the same as the one required in the standard pronunciation task but there is also a lexical decision involved. Performance on the go/no-go pronunciation task could then address the question of whether the increased influence of semantics on lexical decision is due to differences in the response modality (i.e., vocal response vs. button press) or the fact that lexical decision involves a word/non-word discrimination.

[2]The McRae et al. (2005) concrete nouns were selected as stimuli because number of semantic features is available for this set of words.

Table 1 | Descriptive statistics for stimulus characteristics and behavioral data.

Variable ($n = 473$)	*M*	SD
Log frequency (Brysbaert and New, 2009)	2.46	0.62
Number of morphemes	1.23	0.47
Number of letters	5.89	1.94
Number of orthographic neighbors	3.67	4.95
Number of phonological neighbors (Yates, 2005)	7.95	9.69
Orthographic Levenshtein distance (Yarkoni et al., 2008)	2.21	0.92
Phonological Levenshtein distance (Yap and Balota, 2009)	2.05	1.01
Imageability	6.01	0.42
Body object interaction	4.56	1.18
Log number of senses (Miller, 1990)	0.61	0.26
Semantic neighborhood density (ARC; Shaoul and Westbury, 2010)	0.51	0.11
Number of features (McRae et al., 2005)	12.17	3.24
Lexical decision task RTs	600.94	67.90
Lexical decision task accuracy	0.91	0.11
Go/no-go lexical decision task RTs	560.48	69.60
Go/no-go lexical decision task accuracy	0.95	0.09
Pronunciation task RTs	547.64	43.46
Pronunciation task accuracy	0.94	0.07
Progressive demasking task RTs	1429.59	132.00
Progressive demasking task accuracy	0.92	0.10
Semantic classification task RTs	710.05	103.20
Semantic classification task accuracy	0.95	0.08

between orthographic (Yarkoni et al., 2008) and phonological (Yap and Balota, 2009) Levenshtein distance (LD; $r = 0.92$), we used principal component analysis to reduce the two neighborhood size measures and the two LD measures to a neighborhood size (N) and LD component respectively (see Yap et al., 2011).

Semantic variables

Imageability ratings were obtained for 473 of the words from the MRC norms (Coltheart, 1981) and from the norms collected by Cortese and colleagues (Cortese and Fugett, 2004; Schock et al., in press) and Bennett et al. (2011). BOI ratings for 459 of the words were obtained from the Bennett et al. (2011) norms and BOI ratings for the remaining 55 words were collected at the University of Calgary from another separate group of 38 undergraduate students. NF values were taken from the McRae norms, and NS values were log-transformed and were from Miller (1990). Finally, SND values were based on ARC (average radius of co-occurrence) values from Shaoul and Westbury (2010); words whose closest neighbors are *more* similar to them have higher ARC values.

PROCEDURE

In all tasks, testing began with a short series of practice trials with verbal feedback provided by the experimenter. Each of the tasks followed the same general procedure: on each trial, a word was presented in the center of a 20″ monitor controlled by a desktop computer. The LDT, go/no-go LDT, pronunciation task, and SCT were all conducted using E-Prime software (Schneider et al., 2002), while the PDT was run using Windows executable software for the PDT (Dufau et al., 2008).

On each trial in the LDT, participants classified each item as a word or non-word by pressing the far right or far left button on a response box, respectively. On each trial in the go/no-go LDT, participants pressed the far right button to classify an item as a word, and made no response to non-words. On each trial in the pronunciation task, participants read the word aloud into a microphone connected to the response box. Participants' pronunciation responses were recorded with a digital recorder and later coded for accuracy. In the PDT, each item (e.g., TABLE) was presented alternating with a mask (#####), with each trial consisting of a series of repeated cycles. Initially the mask was presented on the screen for 195 ms and the stimulus for 15 ms, and over subsequent cycles the mask presentation time decreased while the stimulus presentation time increased (e.g., in the next cycle the mask would be presented for 180 ms and the stimulus for 30 ms). Participants were instructed to press the space bar on the keyboard as soon as they could determine what the word was. (If no response was made after 2700 ms, the mask disappeared and the stimulus remained on the screen until the participant pressed the space bar.) Participants then typed the word and pressed the enter key to advance to the next trial. Finally, in the SCT, participants classified each word as concrete or abstract by pressing the far right or far left button, respectively.

RESULTS

We excluded trials for which the response was incorrect (7.52% in the LDT, 2.60% in the go/no-go LDT, 3.48% in the pronunciation task, 5.05% in the PDT, and 5.18% in the SCT), as well as responses that were faster than 200 ms or slower than 3000 ms (<1% of trials in all tasks). We also excluded trials for which the RT exceeded 2.5 SD from each participant's mean (2.74, 2.94, 2.55, 2.00, and 2.91% respectively).

There were 473 items for which we had values on each of the lexical and semantic variables examined. For these items, intercorrelations between predictors and dependent measures are presented in **Table 2**. As illustrated in **Table 2**, although there were several significant correlations between the richness variables, most of the correlations were relatively modest (*r*s between 0.05 and 0.28), with the exception of the relationship between NS and SND ($r = 0.49$). However, it should be noted that word frequency is relatively highly correlated with both NS ($r = 0.51$) and SND ($r = 0.77$), suggesting that frequency is driving the NS–SND correlation. Indeed, when one partials out the effect of frequency, the correlation becomes 0.17. Generally, the modest correlations between richness measures suggest that these dimensions do not all tap the same underlying construct, and that semantic richness is multidimensional.

All data were analyzed using R (R Development Core Team, 2011). Linear mixed effects models were fitted to the RT data[3] from each task, using the lme4 package (Bates et al., 2012); *p*-values for

[3]There were too few response errors in most of the tasks to warrant parallel analyses of the accuracy data.

Table 2 | Correlations between predictor variables and dependent measures.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Log frequency	–																
2. Morphemes	−0.20***	–															
3. Letters	−0.46***	0.53***	–														
4. ON	0.42***	−0.26***	−0.68***	–													
5. PN	0.42***	−0.29***	−0.67***	0.79***	–												
6. OLD	−0.50***	0.42***	0.91***	−0.68***	−0.66***	–											
7. PLD	−0.45***	0.46***	0.87***	−0.60***	−0.67***	0.92***	–										
8. Imageability	0.16***	0.01	0.00	−0.08†	−0.05	0.03	0.02	–									
9. BOI	0.33***	0.00	−0.27***	0.32***	0.30***	−0.26***	−0.24***	−0.01	–								
10. NS	0.51***	−0.27***	−0.44***	0.50***	0.48***	−0.48***	−0.43***	−0.06	0.24***	–							
11. SND	0.77***	−0.27***	−0.42***	0.34***	0.36***	−0.45***	−0.38***	0.15**	0.09*	0.49***	–						
12. NF	0.27***	0.05	−0.03	0.06	0.07	−0.04	−0.05	0.28***	0.08†	0.05	0.15**	–					
13. LDT RTs	−0.68***	0.22***	0.47***	−0.39***	−0.37***	0.48***	0.44***	−0.26***	−0.30***	−0.42***	−0.59***	−0.29***	–				
14. G/NG LDT RTs	−0.70***	0.28***	0.55***	−0.43***	−0.41***	0.56***	0.52***	−0.28***	−0.31***	−0.46***	−0.58***	−0.28***	0.85***	–			
15. Pronunciation RTs	−0.58***	0.26***	0.61***	−0.46***	−0.45***	0.62***	0.59***	−0.14**	−0.33***	−0.40***	−0.47***	−0.22***	0.68***	0.72***	–		
16. PDT RTs	−0.58***	0.21***	0.53***	−0.39***	−0.36***	0.49***	0.44***	−0.21***	−0.26***	−0.37***	−0.50***	−0.22***	0.69***	0.72***	0.59***	–	
17. SCT RTs	−0.46***	0.17***	0.36***	−0.26***	−0.21**	0.30***	0.26***	−0.32***	−0.26***	−0.22***	−0.34***	−0.30***	0.64***	0.69***	0.54***	0.59***	–

Log frequency (Brysbaert and New, 2009); morphemes, number of morphemes; letters, number of letters; ON, number of orthographic neighbors; PN, number of phonological neighbors (Yates, 2005); OLD, orthographic Levenshtein distance (Yarkoni et al., 2008); PLD, phonological Levenshtein distance (Yarkoni et al., 2008); CD, log contextual dispersion (Brysbaert and New, 2009); BOI, body object interaction; NS, log number of senses (Miller, 1990); SND, semantic neighborhood density (Shaoul and Westbury, 2010); NF, number of features (McRae et al., 2005); LDT, lexical decision task; G/NG LDT, go/no-go lexical decision task; PDT, progressive demasking task; SCT, semantic classification task.

†$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

fixed effects were computed using the languageR package (Baayen, 2012). The influence of surface, lexical, and semantic richness variables were treated as fixed effects, while participants and items were treated as random variables. The results of these analyses are presented in **Table 3** and **Figure 1**. Clearly, even when lexical variables were controlled for, semantic richness effects were observed in all tasks. At the same time, the pattern of significant richness effects varied across tasks. Imageability and NF were significant predictors in all tasks (although the NF effect was marginal in the PDT). SND was a significant predictor in the standard and go/no-go LDT, while BOI was a significant predictor in all tasks except the PDT. Finally, with the exception of the go/no-go LDT, NS did not significantly predict performance on tasks. Notably, in all cases where significant relationships were observed between latencies and semantic richness measures, the direction of the relationships was the same: relatively greater richness (whether in terms of more senses, or a more highly imageable referent, or more bodily experience, or a denser neighborhood, or more features) was associated with relatively faster latencies.

DISCUSSION

The present study represents the most comprehensive large-scale study of semantic richness effects to date. Using the McRae et al. (2005) concrete words, we investigated the influence of five theoretically influential semantic richness dimensions[4] (imageability, BOI, NS, SND, NF) on lexical processing, using five different word recognition tasks (LDT, go/no-go LDT, speeded pronunciation, PDT, SCT), extending earlier studies (e.g., Duñabeitia et al., 2008; Pexman et al., 2008; Yap et al., 2011) which have considered fewer variables across fewer tasks. It is noteworthy that semantic richness effects could be reliably detected in *all* tasks of lexical processing, even in speeded pronunciation, where meaning does not need to be computed and there is no emphasis on familiarity-based information (see Balota and Chumbley, 1984). More importantly, the present analyses yield a particularly stringent and conservative test of richness effects, given the large number of lexical variables controlled for and the fact that the unique predictive power of each richness dimension was assessed while holding the other four dimensions constant. However, as a counterpoint to the task-*generality* of semantic richness effects, there was also evidence of clear and systematic task-*specificity*. For example, SND effects were reliable only in the standard and go/no-go LDT, while BOI effects were absent in the PDT. We will now consider the implications of these findings in greater detail.

SEMANTIC RICHNESS EFFECTS ARE TASK-GENERAL

In line with previous investigations, the present study provides further evidence that semantic richness effects generalize across disparate word recognition tasks, broadly consistent with the idea that feedback activation from semantics to orthography and phonology is a pervasive aspect of lexical processing (Hino and Lupker, 1996; Pexman and Lupker, 1999; Pexman et al., 2001; Siakaluk et al., 2008). In addition to examining established word recognition measures (i.e., lexical decision, pronunciation, semantic classification), we are the first to assess the influence of multiple richness dimensions on newer paradigms such as the go/no-go LDT and PDT. Researchers have suggested that the latter tasks may help magnify the size of effects of interest by slowing down the recognition process (PDT; Dufau et al., 2008) or by minimizing the role of task-specific decision processes (go/no-go LDT; Perea

[4]Although number of associates (Duñabeitia et al., 2008; Müller et al., 2010) is clearly an important semantic richness variable, including this predictor would have greatly reduced the statistical power of our analyses, given that NoA values were available for only 377 (out of 473) of our stimuli. We did conduct additional analyses examining NoA effects for this subset of words, and found that with the exception of speeded pronunciation, NoA effects were not significant on any task (see also Yap et al., 2011). However, it is possible that these results are specific to the items used in the present study (i.e., concrete nouns).

Table 3 | Estimates for lexical and semantic fixed effects parameters, along with *p*-values based on the *t*-statistic ($n = 473$).

Predictor variable	LDT	G/NG LDT	Pronunciation	PDT	SCT
LEXICAL VARIABLES					
Letters	5.94*	6.47*	7.28***	34.16***	17.87***
Morphemes	−3.52	2.35	−5.86	−16.58	4.28
Log frequency	−44.26***	−41.18***	−17.79***	−66.28***	−32.73***
N component	2.38	2.07	3.11	1.14	−2.56
LD component	2.66	7.90	8.77**	−19.89†	−26.50***
SEMANTIC VARIABLES					
Imageability	−0.26***	−0.33***	−0.08*	−0.35**	−0.39***
BOI	−5.25**	−5.25**	−4.03**	−5.28	−8.30**
NS	−15.71	−20.76*	−5.91	−18.89	−0.22
SND	−82.26*	−68.88*	−21.63	−78.64	−52.61
NF	−2.22**	−1.74*	−1.18**	−2.39†	−2.69**

Letters, number of letters; morphemes, number of morphemes; Log frequency (Brysbaert and New, 2009); N component, composite number of orthographic and phonological neighbors (Yates, 2005); LD component, composite orthographic and phonological Levenshtein distance (Yarkoni et al., 2008); BOI, body–object interaction; NS, log number of senses (Miller, 1990); SND, semantic neighborhood density (Shaoul and Westbury, 2010); NF, number of features (McRae et al., 2005); LDT, lexical decision task; G/NG LDT, go/no-go lexical decision task; PDT, progressive demasking task; SCT, semantic classification task.

$*p < 0.05$; $**p < 0.01$; $***p < 0.001$; $^{\dagger}p < 0.10$.

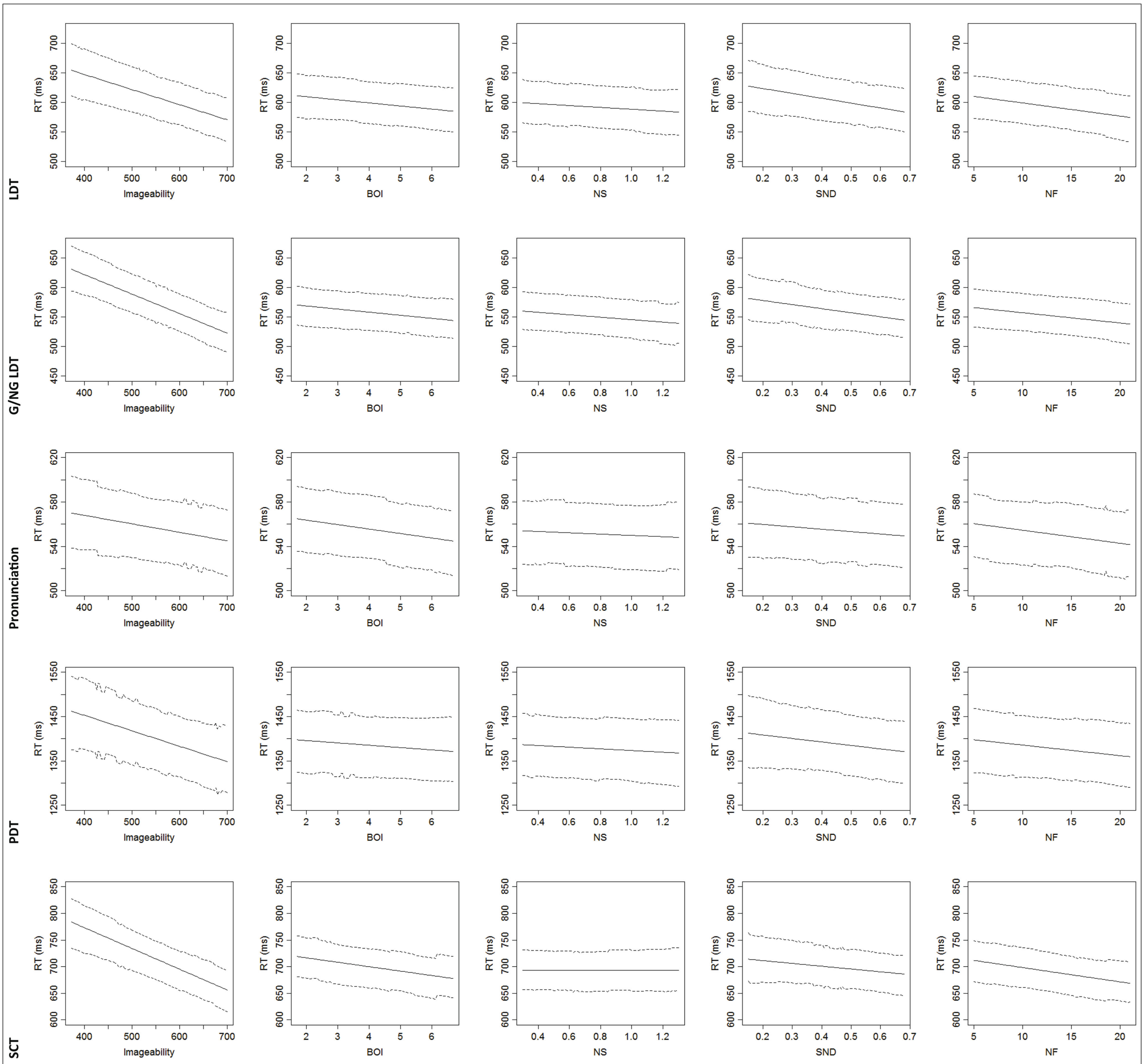


FIGURE 1 | Partial effects plots of semantic richness effects, adjusted for the median value of the other numerical predictors in the model, as a function of task. 95% highest posterior density intervals are provided. Note. BOI, body–object interaction; Senses, log number of senses (Miller, 1990); SND, semantic neighborhood density (Shaoul and Westbury, 2010); Features, number of features (McRae et al., 2005); LDT, lexical decision task; G/NG LDT, go/no-go lexical decision task; PDT, progressive demasking task; SCT, semantic classification task.

et al., 2002). Compared to the standard LDT (see **Table 1**), it is clear that participants were indeed much slower on the PDT, and also faster and more accurate on go/no-go LDT. However, there was no evidence that these two tasks were particularly sensitive to semantic richness effects (cf. Ferrand et al., 2011), compared to the standard LDT. In general, our PDT results are compatible with findings from a recent megastudy that compared performance on lexical decision, pronunciation, and progressive demasking for the same set of 1,482 monosyllabic monomorphemic French words (Ferrand et al., 2011). In that study, Ferrand et al. reported that progressive demasking performance was primarily influenced by perceptual/visual factors such as number of letters, and that the PDT did not provide substantive additional insights beyond those provided by the LDT (but see Carreiras et al., 1997, who reported opposite effects of neighborhood density, a measure of orthographic richness, in the two tasks).

Although we have claimed that semantic richness effects generalize across tasks, we need to carefully qualify this by acknowledging that not *all* effects are reliable in *all* tasks. Specifically, when a large number of lexical and semantic factors were controlled for, the only two richness variables that produced reliable (or borderline reliable) effects on every task were imageability and NF; word recognition was faster for highly imageable words and words for which referents had more features. This suggests that feedback from semantics to orthography during lexical processing is most consistently and robustly mediated by the imaginal and featural aspects of semantic representations. Such a finding is also consistent with frameworks that model semantics via a distributed attractor network (e.g., Plaut and Shallice, 1993), which yields faster settling times for the semantic representations of high-imageability or high-NF words, since these are associated with the activation of more semantic feature units.

It is also noteworthy that BOI effects (Siakaluk et al., 2008) were significant in every task except the PDT; this task-generality is consistent with recent work by Bennett et al. (2011). The pervasiveness of BOI effects strongly supports the idea that the relative availability of sensorimotor information associated with a word contributes to lexical-semantic processing (see also Juhasz et al., 2011), possibly through the recruitment of modality-specific systems (see Hargreaves et al., 2012, for more discussion). Compared to imageability, NF, and BOI, the other richness measures showed more evidence of task-specificity. For example, SND produced reliable effects in the standard and go/no-go LDT, but not in the other three tasks. Likewise, NS did not predict variance on any task except the go/no-go LDT. We will discuss these intriguing between-task dissociations in greater detail in the next section.

SEMANTIC RICHNESS EFFECTS ARE TASK-SPECIFIC

One of the most intriguing aspects of the semantic richness literature is how the strength and even direction of certain semantic richness effects can be systematically and adaptively modulated by the specific demands of a given lexical processing task (see Balota and Yap, 2006). For example, as already observed by Pexman et al. (2008) and Yap et al. (2011), semantic variables collectively account for relatively much more variance in the SCT, compared to other tasks where semantic processing is not the primary basis for a response. At a more fine-grained level, the influence of SND, which provides support for models which structure semantics via lexical co-occurrence (e.g., Shaoul and Westbury, 2010), was evident in the two LDTs, but not in the other three tasks. This replicates the pattern reported by both Pexman et al. (2008) and Yap et al. (2011), and suggests that neighborhood density effects are more reliable in tasks where familiarity-based information is emphasized. Moreover, when a task (i.e., SCT) requires participants to compute the specific meaning of presented words, there might be a trade-off between the facilitatory effects of close neighbors and inhibitory effects of distant neighbors (Mirman and Magnuson, 2006), resulting in null neighborhood effects.

Turning to NS, we were surprised to note that this variable predicted unique variance only in the go/no-go LDT (see **Table 3**), given that previous studies (e.g., Borowsky and Masson, 1996; Hargreaves et al., 2011) reported an ambiguity advantage using the standard LDT. When we examined the zero-order correlations between NS and RT, the relationships were clearly negative (i.e., facilitatory) across all tasks. However, in the mixed effects analyses, when lexical and semantic predictors were more stringently controlled, NS reliably predicted additional unique word recognition variance *only* in the go/no-go LDT, possibly because performance on that task is inherently less noisy and contaminated by task-specific processing demands. We should clarify that ambiguity was operationally defined in the present study using WordNet (Miller, 1990) NS. However, in this metric, the multiple senses of a word may or may not be related to one another.

Rodd et al. (2002) have argued that it is important to distinguish between words with multiple related senses (i.e., polysemes, e.g., TWIST) and words with multiple unrelated meanings (i.e., homonyms, e.g., BARK). Interestingly, Rodd et al. reported that in lexical decision, there is an ambiguity advantage for polysemes (related senses) but an ambiguity *disadvantage* for homonyms (unrelated senses; see also Klepousniotou and Baum, 2007). This is consistent with the idea that the semantic richness of a representation is reinforced by multiple related senses but is undermined by multiple unrelated senses through lexical competition (see Rodd et al., 2002, for more discussion). Because our measure of ambiguity does not distinguish between related and unrelated senses, it is possible that there is a trade-off between the facilitatory effects of related senses and the inhibitory effects of unrelated senses. Of course, it is also possible that this particular metric of NS (counting the number of dictionary senses), despite its objectivity, is a relatively coarse proxy for variability in a word's core meaning (Hoffman et al., 2011).

Hoffman et al. have developed an intriguing new ambiguity measure, *semantic diversity* (SD), using lexical co-occurrence data. Although a full description of their approach is beyond the scope of the present report, they essentially considered all the contexts a word could appear in and computed the similarity between these contexts. A word that can appear in very diverse linguistic contexts (e.g., PART) is considered to be high-SD (i.e., high in ambiguity), while a word that can only occur in a restricted range of contexts (e.g., GASTRIC) is considered low-SD (i.e., low in ambiguity). Interestingly, although the correlation between NS and SD is positive and moderate in strength ($r = 0.41$), words judged explicitly to have few senses nevertheless varied greatly on their SD values (Hoffman et al., 2011), suggesting that this is a more sensitive measure of ambiguity that could be exploited by researchers in future work.

An important limitation of the present work is the item set used. Several of the semantic richness measures are available for only a limited set of items; these tend to be words (nouns) that refer to highly concrete referents. As such, although we examined a large set of these items our results may not necessarily generalize to other item sets. Future research should examine semantic richness effects in other word sets and should extend the study of semantic richness to other word types, such as verbs, in order to gain additional insights about lexical-semantic representation. Future work can also be directed toward exploring non-linear effects of semantic richness dimensions, and possible interactions between these variables.

CONCLUSION

The present study examined the influence of multiple semantic richness dimensions across various tasks of lexical processing. The fact that semantic richness effects could be reliably detected on all tasks attests to their robustness and generality. At the same time, there was ample evidence supporting the multidimensional nature of semantic richness and demonstrating that these dimensions are selectively modulated by task demands. In order to be more fully specified, emerging theories of semantic representation need to take into account this complex interplay between lexical-semantic processes and task-specific mechanisms.

ACKNOWLEDGMENTS

This work was supported in part by a Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery Grant to Penny M. Pexman. We thank Dave Balota, Mike Cortese, Jon Duñabeitia, and Paul Siakaluk for their very useful comments during the preparation of this manuscript.

REFERENCES

Baayen, R. H. (2012). *languageR: Data Sets and Functions with "Analyzing Linguistic Data: A Practical Introduction to Statistics,"* R Package Version 1.4. Vienna: R Foundation for Statistical Computing.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412.

Balota, D. A. (1990). "The role of meaning in word recognition," in *Comprehension Processes in Reading*, eds D. A. Balota, G. B. Flores d'Arcais, and K. Rayner (Hillsdale, NJ: Lawrence Erlbaum Associates), 9–32.

Balota, D. A., and Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 340–357.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S., Spieler, D. H., and Yap, M. J. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. Gen.* 133, 336–345.

Balota, D. A., Ferraro, R. F., and Connor, L. T. (1991). "On the early influence of meaning in word recognition: a review of the literature," in *The Psychology of Word Meanings*, ed. P. J. Schwanenflugel (Hillsdale, NJ: Erlbaum), 187–222.

Balota, D. A., and Yap, M. J. (2006). "Attentional control and flexible lexical processing: explorations of the magic moment of word recognition," *From Inkmarks to Ideas: Current Issues in Lexical Processing*, ed. S. Andrews (New York: Psychology Press), 229–258.

Balota, D. A., Yap, M. J., and Cortese, M. J. (2006). "Visual word recognition: the journey from features to meaning (a travel update)," in *Handbook of Psycholinguistics*, 2nd Edn, eds M. Traxler and M. A. Gernsbacher (Amsterdam: Academic Press), 285–375.

Bates, D. M., Maechler, M., and Dai, B. (2012). *lme4: Linear Mixed-Effect Models Using S4 Classes*, R Package Version 0.999375-42. Vienna: R Foundation for Statistical Computing.

Bennett, S. D. R., Burnett, A. N., Siakaluk, P. D., and Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599 multisyllabic nouns. *Behav. Res. Methods* 43, 1100–1109.

Borowsky, R., and Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 63–85.

Brysbaert, M., and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990.

Buchanan, L., Westbury, C., and Burgess, C. (2001). Characterizing semantic space: neighborhood effects in word recognition. *Psychon. Bull. Rev.* 8, 531–544.

Carreiras, M., and Grainger, J. (2004). Sublexical units and the "front end" of visual word recognition. *Lang. Cogn. Process.* 19, 321–331.

Carreiras, M., Perea, M., and Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: cross-task comparisons. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 857–871.

Chumbley, J. I., and Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Mem. Cognit.* 12, 590–606.

Coltheart, M. (1981). The MRC psycholinguistic database. *Q. J. Exp. Psychol. A* 33, 497–505.

Coltheart, M., Davelaar, E., Jonasson, J. T., and Besner, D. (1977). "Access to the internal lexicon," in *Attention and Performance VI*, ed. S. Dornic (Hillsdale, NJ: Erlbaum), 535–555.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256.

Cortese, M. J., and Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behav. Res. Methods Instrum. Comput.* 36, 384–387.

Cortese, M. J., Simpson, G. B., and Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychon. Bull. Rev.* 4, 226–231.

Dilkina, K., McClelland, J. L., and Plaut, D. C. (2008). A single-system account of semantic and lexical deficits in five semantic dementia patients. *Cogn. Neuropsychol.* 25, 136–164.

Dilkina, K., McClelland, J. L., and Plaut, D. C. (2010). Are there mental lexicons? The role of semantics in lexical decision. *Brain Res.* 1365, 66–81.

Dufau, S., Stevens, M., and Grainger, J. (2008). Windows executable software for the progressive demasking task. *Behav. Res. Methods* 40, 33–37.

Duñabeitia, J. A., Avilés, A., and Carreiras, M. (2008). NoA's Ark: influence of the number of associates in visual word recognition. *Psychon. Bull. Rev.* 15, 1072–1077.

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., and Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: evidence from Chronolex. *Front. Psychol.* 2:306. doi:10.3389/fpsyg.2011.00306

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *J. Exp. Psychol. Gen.* 113, 256–281.

Gordon, B. (1983). Lexical access and lexical decision: mechanisms of frequency sensitivity. *J. Mem. Lang.* 22, 24–44.

Grainger, J., and Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychol. Rev.* 103, 518–565.

Grainger, J., Van Kang, M., and Segui, J. (2001). Cross-modal priming from heterographic homophones. *Mem. Cognit.* 29, 53–61.

Hargreaves, I. S., Leonard, G. A., Pexman, P. M., Pittman, D. J., Siakaluk, P. D., and Goodyear, B. G. (2012). The neural correlates of the body-object interaction effect in semantic processing. *Front. Hum. Neurosci.* 6:22. doi:10.3389/fnhum.2012.00022

Hargreaves, I. S., Pexman, P. M., Pittman, D. J., and Goodyear, B. G. (2011). Tolerating ambiguity: ambiguous words recruit the left inferior frontal gyrus in absence of a behavioral effect. *Exp. Psychol.* 58, 19–30.

Harm, M. W., and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111, 662–720.

Hino, Y., and Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: an alternative to lexical access accounts. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1331–1356.

Hino, Y., and Lupker, S. J. (1998). The effects of word frequency for Japanese kana and kanji words in naming and lexical decision: can the dual-route model save the lexical-selection account? *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1431–1453.

Hino, Y., and Lupker, S. J. (2000). The effects of word frequency and spelling-to-sound regularity in naming with and without preceding lexical decision. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 166–183.

Hino, Y., Lupker, S. J., and Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming and semantic categorization tasks: interactions between orthography, phonology and semantics. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 686–713.

Hino, Y., Pexman, P., and Lupker, S. (2006). Ambiguity and relatedness effects in semantic tasks: are they due to semantic coding? *J. Mem. Lang.* 55, 247–273.

Hodges, J. R., Garrard, P., and Patterson, K. (1998). "Semantic dementia," in *Pick's Disease and Pick Complex*, eds A. Kertesz and D. G. Munoz (New York: Wiley-Liss), 83–104.

Hoffman, P., Rogers, T. T., and Lambon Ralph, M. A. (2011). Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: insights using a novel method to quantify contextual variability in meaning. *J. Cogn. Neurosci.* 23, 2432–2446.

Juhasz, B. J., Yap, M. J., Dicke, J., Taylor, S. C., and Gullick, M. M. (2011). Tangible words are recognized faster: the grounding of meaning in sensory and perceptual systems. *Q. J. Exp. Psychol.* 64, 1683–1691.

Klepousniotou, E., and Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: an advantage for polysemous but not homonymous words. *J. Neurolinguistics* 20, 1–24.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods* 37, 547–559.

Miller, G. A. (1990). WordNet: an on-line lexical database. *Int. J. Lexicography* 3, 235–312.

Mirman, D., and Magnuson, J. S. (2006). "The impact of semantic neighborhood density on semantic access," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, eds R. Sun and N. Miyake (Mahwah, NJ: Erlbaum), 1823–1828.

Müller, O., Duñabeitia, J. A., and Carreiras, M. (2010). Orthographic and associative neighborhood density effects: what is shared, what is different? *Psychophysiology* 47, 455–466.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). *The University of South Florida Word Association, Rhyme, and Word Fragment Norms.* Available at: http://www.usf.edu/FreeAssociation/ [retrieved March 9, 2011].

Paivio, A. (1991). Dual coding theory: retrospect and current status. *Can. J. Psychol.* 45, 255–287.

Perea, M., Rosa, E., and Gómez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Mem. Cognit.* 30, 34–45.

Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychol. Rev.* 114, 273–315.

Pexman, P. M. (in press). "Meaning-level influences on visual word recognition," in *Visual Word Recognition*, ed. J. S. Adelman.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P., Bodner, G., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on word recognition. *Psychon. Bull. Rev.* 15, 161–167.

Pexman, P. M., Hino, Y., and Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1252–1270.

Pexman, P. M., Holyk, G. G., and Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Mem. Cognit.* 31, 842–855.

Pexman, P. M., and Lupker, S. J. (1999). Ambiguity and visual word recognition: can feedback explain both homophone and polysemy effects? *Can. J. Exp. Psychol.* 53, 323–334.

Pexman, P. M., Lupker, S. J., and Jared, D. (2001). Homophone effects in lexical decision. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 139–156.

Piercey, C. D., and Joordens, S. (2000). Turning an advantage into a disadvantage: ambiguity effects in lexical decision versus reading tasks. *Mem. Cognit.* 28, 657–666.

Plaut, D. C., McClelland, J. L., Seidenberg, M., and Patterson, K. E. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115.

Plaut, D. C., and Shallice, T. (1993). Deep dyslexia: a case study of connectionist neuropsychology. *Cogn. Neuropsychol.* 10, 377–500.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422.

Rodd, J., Gaskell, G., and Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: semantic competition in lexical access. *J. Mem. Lang.* 46, 245–266.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime User's Guide.* Pittsburgh: Psychology Software Tools, Inc.

Schock, J., Cortese, M. J., and Khanna, M. M. (in press). Imageability estimates for 3000 disyllabic words. *Behav. Res. Methods.*

Schwanenflugel, P. J. (1991). "Why are abstract concepts hard to understand?" *The Psychology of Word Meanings*, ed P. J. Schwanenflugel (Hillside, NJ: Erlbaum), 223–250.

Seidenberg, M. S., and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychol. Rev.* 96, 523–568.

Shaoul, C., and Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behav. Res. Methods* 42, 393–413.

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., and Sears, C. R. (2008). Evidence for the activation of sensorimotor information during visual word recognition: the body-object interaction effect. *Cognition* 106, 433–443.

Strain, E., Patterson, K., and Seidenberg, M. S. (1995). Semantic effects in single-word naming. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 1140–1154.

Tillotson, S. M., Siakaluk, P. D., and Pexman, P. M. (2008). Body-object interaction ratings for 1,618 monosyllabic nouns. *Behav. Res. Methods* 40, 1075–1078.

Wellsby, M., Siakaluk, P. D., Owen, W. J., and Pexman, P. M. (2011). Embodied semantic processing: the body-object interaction effect in a non-manual task. *Lang. Cogn.* 3, 1–14.

Yap, M. J., and Balota, D. A. (2009). Visual word recognition of multisyllabic words. *J. Mem. Lang.* 60, 502–529.

Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.

Yarkoni, T., Balota, D. A., and Yap, M. J. (2008). Beyond Coltheart's N: a new measure of orthographic similarity. *Psychon. Bull. Rev.* 15, 971–979.

Yates, M. (2005). Phonological neighbors speed visual word processing: evidence from multiple tasks. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1385–1397.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 05 January 2012; accepted: 15 March 2012; published online: 17 April 2012.

*Citation: Yap MJ, Pexman PM, Wellsby M, Hargreaves IS and Huff MJ (2012) An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. Front. Hum. Neurosci. **6**:72. doi: 10.3389/fnhum.2012.00072*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 14 February 2012
doi: 10.3389/fnhum.2012.00011

The time course of semantic richness effects in visual word recognition

Milena Rabovsky[1,2]*, Werner Sommer[1,2] and Rasha Abdel Rahman[1]

[1] *Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany*
[2] *Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany*

Edited by:
Penny M. Pexman, University of Calgary, Canada

Reviewed by:
Ian S. Hargreaves, University of Calgary, Canada
Ben Amsel, University of California, San Diego, USA

****Correspondence:***
Milena Rabovsky, Department of Psychology, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany.
e-mail: milena.rabovsky@hu-berlin.de

The richness of semantic representations associated with individual words has emerged as an important variable in reading. In the present study we contrasted different measures of semantic richness and explored the time course of their influences during visual word processing as reflected in event-related brain potentials (ERPs). ERPs were recorded while participants performed a lexical decision task on visually presented words and pseudowords. For word stimuli, we orthogonally manipulated two frequently employed measures of semantic richness: the number of semantic features generated in feature-listing tasks and the number of associates based on free association norms. We did not find any influence of the number of associates. In contrast, the number of semantic features modulated ERP amplitudes at central sites starting at about 190 ms, as well as during the later N400 component over centro-parietal regions (300–500 ms). Thus, initial access to semantic representations of single words is fast and word meaning continues to modulate processing later on during reading.

Keywords: word meaning, semantic richness, visual word recognition, ERPs, N400

INTRODUCTION

Extracting meaning from words and texts is the ultimate goal of written language comprehension. Yet, semantic representations and the mechanisms underlying semantic processing remain elusive. Accordingly, many models of visual word recognition, even though assuming a role for semantic representations, have restricted explicit computational implementations to orthographic and phonological processes (Seidenberg and McClelland, 1989; Plaut et al., 1996; Coltheart et al., 2001; but see Harm and Seidenberg, 2004). One way to approach the issue how meaning is represented and retrieved during reading is to examine influences of semantic richness, that is, how differences between words in the amount of associated semantic information modulate word processing (Buchanan et al., 2001; Locker et al., 2003; Yates et al., 2003; Balota et al., 2004; Adelman et al., 2006; Pexman et al., 2007; Duñabeitia et al., 2008; Mirman and Magnusson, 2008; Pexman et al., 2008; Grondin et al., 2009; Yap et al., 2011). Investigating, which of the various measures proposed to quantify semantic richness influence word processing, and at what point in time these influences take place (absolutely as well as in relation to other lexical processes), helps to specify the nature of semantic representation and to come closer to an understanding of the processes taking place when meaning is extracted from print. As yet, evidence on the time course of semantic richness effects during reading is scarce and inconsistent (Kounios et al., 2009; Müller et al., 2010; Amsel, 2011; see below). Exploiting the high temporal resolution provided by event-related brain potentials (ERPs), we aimed to further clarify this issue.

Measures of the richness of semantic representations have been based, for example, on word co-occurrences in text corpora (Buchanan et al., 2001), contextual dispersion across different content areas (Adelman et al., 2006), and the number of semantic features generated in feature-listing tasks (McRae et al., 2005). However, Pexman et al. (2008) reported different patterns of contributions for these different measures of semantic richness to the latencies of lexical decisions and semantic categorizations. Hence, semantic richness seems not to be a unitary phenomenon and distinct mechanisms may underlie the influences of different facets of this variable. Here, we directly contrasted two important measures of the richness of semantic representations, the number of semantic features and associates, as described below.

Many influential models and theories assume semantic features to play a crucial role in meaning representation (e.g., Collins and Loftus, 1975; Plaut and Shallice, 1993; Harm and Seidenberg, 2004); the number of semantic features determining a word's meaning thus seems to be a key indicator of semantic richness. Indeed, words with many semantic features (e.g., desk) are processed faster in lexical decision and semantic categorization tasks than words with fewer features (e.g., cork; Pexman et al., 2002, 2003; Grondin et al., 2009). In the present study the number of semantic features was manipulated based on the elaborate norms by McRae et al. (2005) where more than 700 participants had listed semantic features for 541 concrete words (e.g., mouse—"is small," "has legs," etc.).

Another relevant and frequently used measure of semantic richness is the number of different first associations generated across participants in free-association tasks (Nelson et al., 2004). In the study by Nelson et al. (2004), participants produced the first word that came to their mind upon hearing a specific cue word. Subsequently, the number of different first associations to this cue word was counted, excluding idiosyncratic associations (i.e., associated words produced by a single participant only). Based on the assumption that every first free association to a cue word is an associate in semantic memory, the number of different first free associations has been referred to as the number of associates. For simplicity, we will use the term number of associates throughout the manuscript. Words with many associates have been found to be processed faster than words with few associates in tasks such as lexical decision, semantic categorization, reading aloud, perceptual identification, and online sentence reading (Buchanan et al., 2001; Pexman et al., 2007; Duñabeitia et al., 2008). However, it is important to note that a recent study did not find any influences of the number of associates when controlling for other lexical and semantic variables such as e.g., the number of features (Yap et al., 2011). Thus, the available evidence concerning independent influences of the number of associates is controversial.

An important step toward understanding the mechanisms underlying the extraction of meaning from print, and ultimately specifying these mechanisms in explicit computational models, is to investigate the time course of influences of semantic richness, that is, at which time it affects word processing and how the timing of semantic richness effects relates to influences of other lexical variables. A number of studies have investigated the temporal dynamics of different aspects of semantic processing, and recent evidence suggests that lexical semantic access may start as early as within the first 200 ms of word processing (e.g., Skrandies, 1998; Hauk et al., 2006; Penolazzi et al., 2007; Kiefer et al., 2008; Dambacher et al., 2009; Pulvermüller et al., 2009; Segalowitz and Zheng, 2009; Kiefer and Pulvermüller, 2011; Rabovsky et al., 2011a). However, evidence on the time course of the above-mentioned effects of semantic richness is restricted to three recent ERP studies, two focusing on semantic features (Kounios et al., 2009; Amsel, 2011) and the other on associates (Müller et al., 2010).

Kounios et al. (2009) aimed at a graded manipulation of semantic richness, ranging from abstract words with presumably rather poor representations in semantic space (Paivio, 1986; Plaut and Shallice, 1993), over concrete words with few semantic features, to concrete words with many semantic features (cf. McRae et al., 2005). In this study, participants were presented with word pairs, which could be semantically related or unrelated. Whereas no responses were to be given to the first words of the pairs (the experimental items), the second words of the pairs had to be judged for their relatedness to the preceding word. Significant effects of semantic richness were obtained in the ERPs between 200 and 800 ms after stimulus onset, but the theoretically most extreme comparison (between abstract words and concrete words with many semantic features) was not significant in any segment. Furthermore, ERP amplitudes did not show a monotonic ordering according to the presumably graded variation of semantic richness. In line with previous evidence (Kounios and Holcomb, 1994; West and Holcomb, 2000), N400 amplitudes to both types of concrete words were larger as compared to abstract words; however, concrete words with few features unexpectedly tended to elicit larger N400 amplitudes than those with many semantic features. Because this result is at variance with predictions, Kounios et al. (2009) concluded that semantic richness either has a non-monotonic effect on neural activity or that, alternatively, their manipulation of semantic richness was confounded with some other factor.

In a study by Amsel (2011), participants read words silently and subsequently made two judgments, at first about the extent to which the word elicited mental imagery and then concerning the extent to which this imagery was based on specific personal memories. In contrast to Kounios et al. (2009), Amsel indeed found more negative amplitudes for words with more semantic features starting at about 320 ms. Furthermore, this study reported significant influences of the number of features between 200 and 300 ms, and an additional short-lived effect already at 120 ms. Thus, available evidence on influences of the number of features is rather mixed concerning both timing and direction.

Müller et al. (2010) manipulated the number of associates in a lexical decision task and observed larger N400 amplitudes for words with more associates. This seems in line with the finding of enhanced negativity for words with many features at about 320 ms (Amsel, 2011) but in disagreement with the observation of Kounios et al. (2009) that concrete words with fewer semantic features elicited larger N400 amplitudes. It seems important to note that part of the inconsistency may be due to the use of different measures of semantic richness, as the relation of the different variables is currently not clear. Furthermore, in line with theories assuming a feature-based organization of semantic memory (Plaut and Shallice, 1993; Harm and Seidenberg, 2004; McRae, 2004), recent evidence suggests that the number of associates may not yield independent contributions to semantic richness effects when other relevant variables are controlled for (Yap et al., 2011). In sum, the evidence concerning the temporal evolvement of semantic richness effects during reading is far from conclusive.

It seems especially interesting to pinpoint the moment when semantic richness effects first arise during reading, and to determine the temporal delay between initial access to form-related lexical information and the activation of the corresponding semantic representations. Access to orthographic representations has been proposed to be reflected in the left-lateralized N1 component of the ERP peaking at about 160 ms, presumably corresponding to hemodynamic activation in an area within the left fusiform gyrus assumed to be specialized in visual word form processing (e.g., McCandliss et al., 2003; Maurer et al., 2005; Brem et al., 2009). As activation of orthographic representations may enable the retrieval of the corresponding semantic information, the subsequent components (P2 and N2) seem to be interesting candidates for initial influences of semantic richness. Indeed, these components seem to be modulated, for example, by semantic context (van den Brink et al., 2001; Kandhadai and Federmeier, 2010; Barber et al., 2011).

Aiming to specify the temporal relationship between word form and meaning processing, we compared the onset of semantic richness effects with the onset of lexicality effects: as pseudowords do not match any pre-existing visual word form representation, ERP differences between words and pseudowords may already arise at the level of orthographic processing, preceding possible effects at the semantic level.

In sum, we recorded ERPs while participants performed visual lexical decisions. Within the word stimuli, we orthogonally manipulated two prominent measures of semantic richness, namely the number of semantic features (McRae et al., 2005) and associates (Nelson et al., 2004) to assess independent contributions of each variable. Based on the evidence for early lexical semantic modulations described above, we examined influences of semantic richness not only on the N400 component, which has often been related to semantic processes (Kutas and Federmeier, 2011), but also on earlier ERP components.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-four native English speakers (from Australia, New Zealand, the UK, and the USA) were paid 7€ per hour to participate in the study. Half of them were male; their age ranged from 19 to 32 ($M = 25$) years. All participants had normal or corrected-to-normal vision, 20 were right-handed. Written informed consent was obtained before the experiment.

MATERIALS

Stimuli were 160 concrete English nouns (40 per condition combination) and 160 pseudowords. Within the word stimuli, the number of semantic features (McRae et al., 2005) and the number of associates (Nelson et al., 2004) were orthogonally manipulated across two levels each. The four condition combinations did not differ in familiarity, concreteness, word length, word frequency, bigram frequency, or in the number of orthographic neighbors, phonological neighbors, phonemes, and syllables (all $Fs < 1$; please see **Table 1**). We report the measures for number of associates and concreteness based on Nelson et al. (2004). The word frequency values represent log-transformed frequencies based on the HAL corpus (Lund and Burgess, 1996), according to the English Lexicon Project (ELP; Balota et al., 2007). Values for the remaining word characteristics are taken from McRae et al. (2005). Pseudowords were constructed by recombining the letters of the word stimuli (e.g., "osnop" from "spoon"). They were orthographically less typical than the words as indicated by bigram and trigram frequency values retrieved from the ELP (Balota et al., 2007).

PROCEDURE

Participants were seated in a dimly lit, sound-attenuated, and electrically shielded chamber. Stimuli were presented in black in 24 point Arial font on a light blue screen, 1 m in front of the participants. Each trial began with a fixation cross, presented for 1.5 s, and followed by a letter string, terminated by the response or after 3 s. The next trial started immediately thereafter, with a constant response-stimulus interval of 1.5 s. Participants were instructed to restrict eye blinks to the periods during which the fixation cross was visible. Participants were to indicate as fast and accurately as possible whether the letter string was a word or not by pressing a button to the left or right with the corresponding index finger. Response hand-to-stimulus assignments were counterbalanced across participants. Stimuli were presented in a different random order for each participant. In all, the experiment comprised 320 trials, separated by seven short breaks.

Table 1 | Stimulus characteristics.

Features	Many		Few	
Associates	Many	Few	Many	Few
N° features	16.08	16.08	9.25	9.25
N° associates	18.03	9.20	18.05	9.20
Familiarity	6.24	6.40	6.16	6.01
Concreteness	6.07	6.08	5.99	6.01
Length (n° letters)	5.53	5.33	5.35	5.15
Word frequency	8.31	8.20	8.28	8.26
Bigram frequency	3825	3507	3334	3500
N° orth. neighbors	5.63	7.58	6.55	6.23
N° phon. neighbors	13.90	14.80	13.75	15.10
N° phonemes	4.45	4.30	4.50	4.18
N° syllables	1.63	1.60	1.58	1.55

EEG RECORDING AND ANALYSIS

The EEG was recorded with Ag/AgCl electrodes from 62 scalp sites according to the extended 10–20 system, and referenced to the left mastoid. Electrode impedance was kept below 5 kΩ. Bandpass of amplifiers (Brainamps) was 0.032–70 Hz, and sampling rate was 500 Hz. Offline, the EEG was transformed to average reference. Eye blink artifacts were removed with a spatio-temporal dipole modeling procedure using BESA software (Berg and Scherg, 1991). After applying a 30 Hz low pass filter, the continuous EEG was segmented into epochs of 1 s, with a 200 ms pre-stimulus baseline. Trials with remaining artifacts and with incorrect or missing responses were discarded.

Early parts of the ERP waves were segmented based on measures of global map dissimilarity (GMD; Lehmann and Skrandies, 1980; Brandeis et al., 1992). GMD values reflect dissimilarities of topographies across adjacent time points so that GMD peaks indicate transitions between periods of relative topographical stability, which indicate ongoing processes in similar brain areas (Lehmann and Skrandies, 1980). GMD values based on ERPs averaged across all experimental conditions peaked at 80, 130, 180, and 240 ms (see **Figure 1D**). These moments of transition were taken as borders for ERP segmentation; thus, subsequent analyses focused on segments between 80–130, 130–180, and 180–240 ms. As word processing continues during later parts of the ERP, transitions between brain states as indicated by GMD measures become less clear-cut (see **Figure 1D**). Thus, a further segment between 300 and 500 ms, corresponding to the N400 component, was chosen based on the literature (e.g., Kutas and Federmeier, 2011).

Amplitudes were averaged within these selected epochs. We first analyzed amplitudes of global field power (GFP; Lehmann and Skrandies, 1980), reflecting the average activity across all electrodes (see **Figures 1A–C**). By providing a global measure of activity across the scalp, GFP analyses diminish the risk of obtaining false positive results, which may be entailed by focusing on a few electrode sites only. Furthermore, for each segment, these global analyses were complemented by analyses focusing on electrode sites at relevant regions of interest (ROIs). For early components with sharp and clearly localized peaks, these complementary analyses focused on electrodes with maximal amplitudes (averaged across all experimental conditions) and their contralateral counterparts: PO7/PO8 for the segment between 80 and 130 ms, corresponding to the P1 component, and PO9/PO10 for

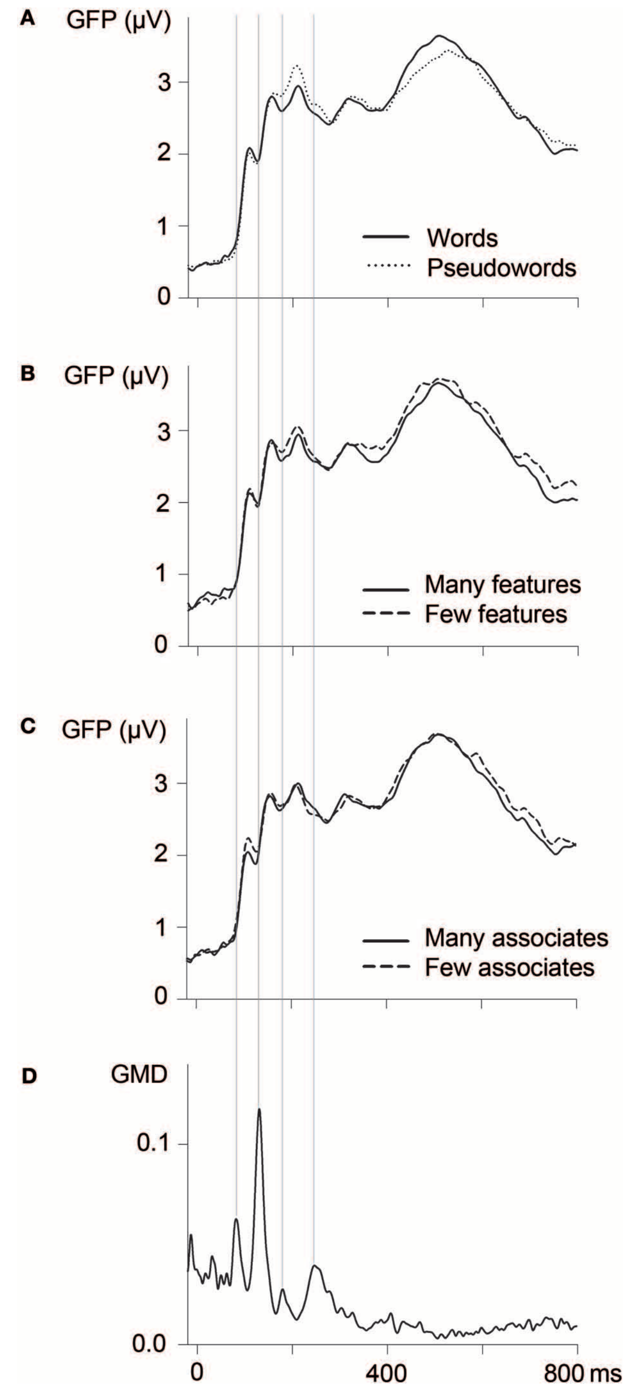


FIGURE 1 | Global field power (GFP; $n = 24$; 62 channels) as a function of lexicality **(A)**, the number of semantic features **(B)**, and the number of associates **(C)**. Vertical lines indicate borders of ERP segmentation, based on measures of global map dissimilarity (GMD; **D**).

the segment between 130 and 180 ms, corresponding to the N1 component. The segment between 180 and 240 ms corresponds to the P2/N2 complex, with a negative maximum at posterior sites and a positive maximum over the vertex, so that we analyzed ERPs both at posterior (PO9/PO10) and at central (C1, Cz, CPz, C2) sites. For the N400 component, which has a broad maximum over centro-parietal regions (Kutas and Federmeier, 2011), we analyzed a larger electrode cluster (C1, Cz, C2, CP1, CPz, CP2, P1, Pz, P2) between 300 and 500 ms.

Even with painstaking controls of perceptual factors the use of different words for different experimental conditions always induces the risk that obtained effects may be due to sensory confounds. Especially when focusing on early visual components, this is a serious issue. In order to control for such possible confounds, pseudowords were categorized according to the semantic richness condition of the words they were derived from. As pseudowords were constructed by recombining the letters of the word stimuli (e.g., "osnop" from "spoon"), they were identical to their base words in terms of the basic visual features contained in the letters, but differed from their base words in that they did not convey meaning. Hence, we applied the ERP analyses run on the word stimuli in an analogous way to the pseudoword stimuli. Thus, obtaining semantic richness effects for the words but not for the pseudowords provides evidence against an interpretation of the obtained semantic richness effects in terms of a sensory confound.

In addition to semantic richness effects, we also analyzed differences between words and pseudowords. As effects of semantic richness can be analyzed only within words, data were thus submitted to two types of ANOVAs. One focused on word stimuli only and included the factors Features (many vs. few) and Associates (many vs. few), while the other was applied to all stimuli and included the factor Lexicality (words vs. pseudowords). *Post-hoc* tests were Bonferroni-corrected.

In an attempt to capture the temporal delay between the influences of lexicality and semantic richness, we calculated t-tests for each time point between GFP amplitudes to words vs. pseudowords (lexicality effect), as well as between GFP amplitudes to words with many vs. few semantic features (feature effect). Onsets were defined as the points in time when an effect first started to be significant (for $df = 23$ $p < 0.05$ if $t > 2.069$) over five successive sampling points.

In order to compare the scalp topographies of the effects across time windows, difference waves were scaled to the individual GFP within the relevant time windows for each participant; that is, the amplitude at each electrode was divided by GFP. This was done in order to omit differences in overall amplitude since only the shape of the distributions was to be compared.

RESULTS

PERFORMANCE

For each subject, we excluded trials deviating from the subject's mean RT by more than 2 SDs. Response latencies were significantly shorter for words ($M = 682$ ms) than for pseudowords ($M = 759$ ms), $F1(1{,}23) = 20.25$, $p < 0.001$, $F2(1{,}318) = 123.1$, $p < 0.001$, but were not affected by the semantic richness factors, $F1s < 1$, $F2s < 1$. Analyses of error rates did not reveal significant effects, with overall very high accuracy for both words ($M = 1.8\%$) and pseudowords ($M = 2.0\%$).

ELECTROPHYSIOLOGY

GFP amplitudes as a function of lexicality, the number of features, and the number of associates are depicted in **Figures 1A–C**.

F-values and significance levels for analyses of GFP amplitudes and for the complementary analyses of ERP amplitudes at specific ROIs (see below for details) are reported in **Table 2**.

Between 80 and 130 ms, in the segment corresponding to the P1 component, neither analyses of GFP amplitudes nor analyses focusing on the electrodes with maximal P1 amplitudes (PO7/PO8) showed significant effects.

Between 130 and 180 ms, corresponding to the N1 time window, analyses of electrodes with maximal N1 amplitudes (PO9/PO10) revealed a significant influence of Electrode Site, $F(1,23) = 16.16$, $p = 0.001$, indicating left-lateralization as typically observed for the N1 to visual words (e.g., McCandliss et al., 2003). In addition, there was a significant influence of lexicality (with a corresponding trend in the GFP analysis), indicating larger N1 amplitudes for pseudowords than words (see **Figures 1 and 2**).

In the segment between 180 and 240 ms, corresponding to the P2/N2 complex, analyses of GFP amplitudes showed continued influences of lexicality, which could be confirmed at the posterior sites PO9 and PO10, $F(1,23) = 16.64$, $p < 0.001$ ($F < 1$ at C1, Cz, CPz, C2). Comparison of topographical distributions of lexicality effects between the earlier (130–180 ms) and later (180–240 ms) segment revealed no significant difference, $F(1,23) = 1.17$, $p = 0.32$. Importantly, during the segment between 180 and 240 ms, GFP analyses also revealed significant influences of the number of features (see **Figure 1 and Table 2**). Feature effects could be confirmed over the vertex, $F(1,23) = 4.31$, $p < 0.05$ at C1, Cz, CPz, C2 (please see **Figure 3**), but were not significant over posterior areas, $F(1,23) = 1.82$, $p = 0.19$.

Running t-tests on GFP amplitudes (see Methods) indicated lexicality effects to start at 164 ms while feature effects arose at 190 ms, suggesting that semantic features are activated only about 20–30 ms after form-related properties during word reading.

During the N400 segment, GFP analyses did not reveal significant results. However, the complementary analysis focusing on centro-parietal sites (C1, Cz, C2, CP1, CPz, CP2, P1, Pz, P2), providing increased sensitivity, revealed a significant influence of the number of features, $F(1,23) = 4.83$, $p < 0.05$. As can be seen in **Figure 3**, amplitudes were more negative for words with more semantic features (and even though not statistically reliable, this tendency is also observable in GFP amplitudes depicted in **Figure 1**). A comparison of topographical

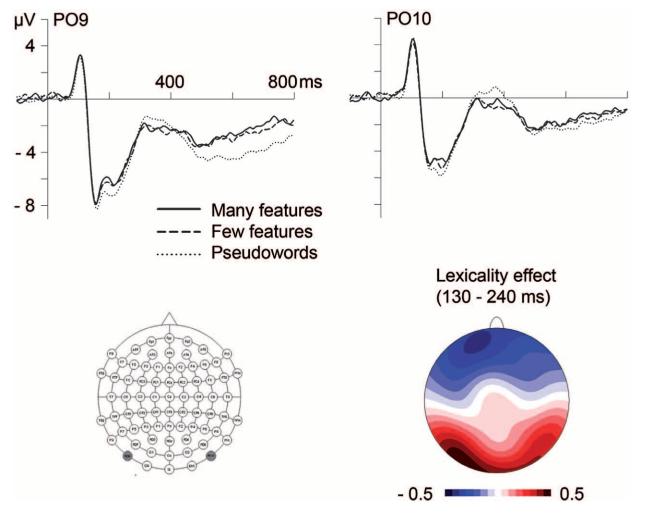


FIGURE 2 | Top. Influences of lexicality and the number of semantic features on event-related brain potentials at posterior electrode sites. **Bottom**. At left is a map of electrode locations with the depicted sites PO9 and PO10 highlighted in dark gray. To the right is the topographical distribution of the lexicality effect (words minus pseudowords) between 130 and 240 ms.

Table 2 | *F*-values and significance levels from analyses of variance on GFP amplitudes and on ERP amplitudes at relevant electrode sites (ROIs; see Methods and Results for details).

Source	*df*	Time segments							
		P1 (80–130)		N1 (130–180)		P2/N2 (180–240)		N400 (300–500)	
		GFP	ROI	GFP	ROI	GFP	ROI	GFP	ROI
WORDS									
Features	1, 23					4.82*	4.31*		4.83*
Associates	1, 23								
Features × Associates	1, 23								
PSEUDOWORDS									
Features	1, 23								
Associates	1, 23								
Features × Associatets	1, 23								
ALL STIMULI									
Lexicality	1, 23			3.42 (*)	5.81*	18.82***	16.64***		

Words were analyzed to examine influences of semantic richness; pseudowords differing in the semantic richness of the words they were derived from were analyzed analogously to control for possible contributions of sensory confounds. Analyses on all stimuli examined lexicality effects.
****p < 0.001; **p < 0.01; *p < 0.05; (*) p = 0.077.*

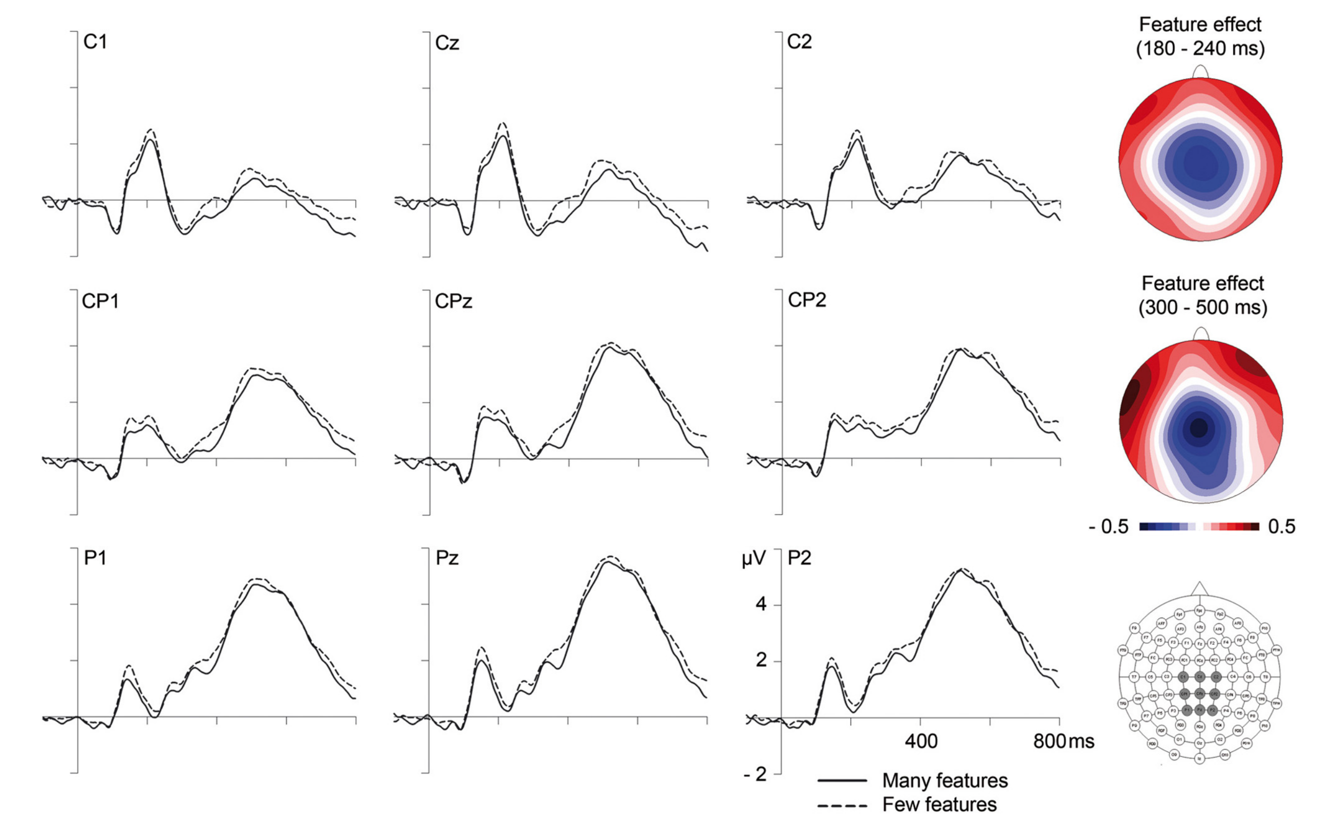


FIGURE 3 | Influences of the number of semantic features on event-related brain potentials at centro-parietal electrode sites. On the right are topographical distributions of feature effects (many minus few semantic features) between 180 and 240 (top) and between 300 and 500 ms (middle), as well as a map of electrode locations with the depicted sites highlighted in dark gray.

distributions of semantic feature effects between the earlier (180–240) and later (300–500) time windows revealed no significant difference ($F < 1$).

Analogous analyses of ERPs to pseudowords categorized according to the semantic richness of their base words did not reveal any significant differences (see **Table 2**).

DISCUSSION

The present study investigated the time course of semantic richness effects during visual word recognition by means of ERPs. We focused on two different measures of semantic richness, namely the number of semantic features (McRae et al., 2005) and the number of associates (Nelson et al., 2004). Of primary interest were, whether, and how these measures contribute to semantic richness effects during word reading, and to disentangle their relative contributions. In addition, we related the onset of semantic influences to the onset of lexicality effects in order to obtain relative temporal information on the time course of word form and meaning access. The number of semantic features modulated ERP amplitudes starting at about 190 ms, shortly after the onset of lexicality effects during the N1 segment at about 164 ms. Later on, in the N400 segment, the number of semantic features enhanced negative amplitudes at centro-parietal sites. In contrast, we did not find any influence of the number of associates. We will detail and discuss these findings below.

EARLY ERP COMPONENTS

The first ERP component found to be modulated was the posterior left-lateralized N1, presumably reflecting visual word form processing within the fusiform gyrus (McCandliss et al., 2003; Maurer et al., 2005; Brem et al., 2009; see **Figure 2**). Mean amplitudes at posterior sites in the N1 segment (130–180 ms) were modulated by lexicality, with larger amplitudes for pseudowords than for words. This lexicality effect is in line with PET studies showing stronger left fusiform activations for pseudowords than for real words (Brunswick et al., 1999; Fiez et al., 1999). More generally, it fits well with the assumption that the left-lateralized N1 component in reading indicates orthographic activation in the visual word form area (e.g., McCandliss et al., 2003), which seems to be hierarchically organized to code orthographic representations of increasing complexity from individual letters over bigrams and trigrams to whole words (Vinckier et al., 2007). Please note that due to our orthographically untypical pseudowords, the ERP difference between words and pseudowords obtained here may arise at an orthographic locus beneath the whole word level.

Shortly after the N1, an effect of the number of semantic features was observed during the P2/N2 segment (180–240 ms) while the lexicality effect continued (see **Figures 1,2,3**). Thus, semantic access seems to start quickly, within the first 200 ms of reading, in line with recent evidence as discussed in the introduction (Skrandies, 1998; Hauk et al., 2006; Penolazzi et al., 2007; Kiefer et al., 2008; Dambacher et al., 2009; Pulvermüller et al., 2009; Segalowitz and Zheng, 2009; Kiefer and Pulvermüller, 2011; Rabovsky et al., 2011a). Running t-tests on GFP amplitudes (see Methods) indicated lexicality effects to start at 164 ms, while feature effects arose at 190 ms.

As noted above, because the pseudowords were orthographically untypical, our lexicality effect may arise at an orthographic level below access to whole word representations already. Therefore, the temporal delay between access to whole word representations and semantics may be even shorter than the observed 20–30 ms. This suggests that word form activation initially precedes the activation of semantic features by no more than 20–30 ms, and that form-related and semantic properties are subsequently processed in parallel. These results are incompatible with theories assuming discrete and modular processing stages in reading, where processing at a lower orthographic level needs to be completed in order to enable the activation of higher-level semantic representations. Instead, our data support partial information transmission and temporal overlap between processes at different levels of representations in reading (Seidenberg and McClelland, 1989; Plaut and Shallice, 1993; Harm and Seidenberg, 2004).

N400 COMPONENT

As to be expected, the number of semantic word features also affected ERP amplitudes during the time window of the N400 component, which has been related to semantic processing (Kutas and Hillyard, 1980; Kutas and Federmeier, 2000, 2011). Words with many semantic features elicited larger N400 amplitudes at centro-parietal sites than words with fewer features (see **Figure 3**), in line with Amsel (2011), reporting an enhanced negativity for words with many features from about 320 ms onwards. Furthermore, our results fit well with findings that concrete words—considered to contain richer semantic representations—produce larger N400 amplitudes than abstract words (Kounios and Holcomb, 1994; West and Holcomb, 2000; Kounios et al., 2009), and that newly learned objects and their written names elicit larger N400 amplitudes when they are associated with in-depth as compared to minimal semantic information (Abdel Rahman and Sommer, 2008; Rabovsky et al., 2011a).

On the other hand, our finding of enhanced N400 amplitudes for words with more semantic features is at variance with the results of Kounios et al. (2009) who observed a trend for larger N400 amplitudes for words with fewer semantic features. However, the authors themselves found it surprising that their semantic feature effect was in the opposite direction as the commonly observed concreteness effect that they had also replicated in their study, and accordingly discussed the possibility that their semantic richness manipulation might have been confounded with some other factor. Clearly, further research seems desirable. In any case, the present result of enhanced N400 amplitudes for words with many semantic features is in line with feature effects reported by Amsel (2011) as well as ERP effects of concreteness (Kounios and Holcomb, 1994; West and Holcomb, 2000; Kounios et al., 2009) and the amount of newly acquired semantic information (Abdel Rahman and Sommer, 2008; Rabovsky et al., 2011a).

In principle, our finding of larger N400 amplitudes for words with many semantic features would seem to be also compatible with the enhanced negativity in the N400 window for words with more associates reported by Müller et al. (2010). However, although both the number of features and associates measure some facet of semantic richness, and would thus be expected to elicit similar influences, it seems somewhat surprising that we only found effects of the number of features and no influences of the number of associates. On the other hand, our results are in line with Yap et al. (2011) who did find independent influences of the number of features but not the number of associates when controlling for other relevant semantic and lexical variables. Notably, Müller et al. (2010) focused on the number of associates, but did not control for the number of features. As the number of features and associates are positively correlated if not intentionally disentangled as done here, it is possible that the effect of number of associates reported by Müller et al. was at least partly due to the number of semantic features. Furthermore, their stimuli with high and low numbers of associates also differed in imageability, with significantly higher imageability values for words with more associates (see p. 458 and Table 1 in Müller et al., 2010); N400 amplitude enhancements as in the study by Müller et al. have also been found for words with high imageability (Kounios and Holcomb, 1994; Holcomb et al., 1999; West and Holcomb, 2000; Swaab et al., 2002). On the other hand, the discrepancy may also be due to the manipulation of the number of associates being rather modest in our study (mean difference of nine associates between the groups) as compared to the manipulation by Müller and colleagues (mean difference of 24 associates).

Another possibly relevant factor is that the present study employed a lexical decision task with orthographically rather untypical pseudowords; hence, semantic access presumably contributed little to successful task performance. It has been repeatedly shown that semantic influences on word processing depend on task demands (West and Holcomb, 2000; Pexman et al., 2008). The semantic influences elicited in our task were presumably restricted to those influences, which take place automatically when presented with a visual word, and were not induced by task demands and intentional semantic processing. This may also be responsible for the absence of behavioral facilitation for words with richer semantic representations[1], which might have been expected based on previous evidence (Buchanan et al., 2001; Pexman et al., 2007; Duñabeitia et al., 2008; Grondin et al., 2009).

For these reasons, even though it is an interesting topic whether the organization of the semantic system is based on semantic features, associations, or both (Lucas, 2000; Hutchison, 2003; Yee et al., 2009), we would not want to base too strong of a claim on the absence of ERP effects of the number of associates. Still, it seems interesting to note that our findings converge with

[1]While in our performance data, neither ANOVAs nor multiple regression analyses did reveal significant effects of the number of features or associates, multiple regression analyses (but not ANOVAs) on RTs to the same stimuli as retrieved from the ELP showed the expected facilitating influence of the number of features ($p < 0.05$; one-sided) but not associates ($p = 0.17$; one-sided).

Yap et al. (2011) in suggesting that the number of associates may not independently contribute to semantic richness effects on (concrete) word processing when other relevant semantic and lexical variables are controlled for. Notably, the amount of semantic features modulated the ERPs in both early and later time windows in spite of the above-mentioned constraints, in line with automatic task-independent activation of semantic features during reading.

At present it seems difficult to draw clear conclusions concerning the functional basis of the N400 modulation. Possibly it reflects some continued reverberation and settling still related to lexical semantic access. On the other hand, the N400 effect may also reflect additional post-lexical semantic processing or implicit memory formation (see e.g., Schott et al., 2002; Rabovsky et al., 2011b). In any case, the present observation of feature effects being present already between 180 and 240 ms seems to converge with earlier suggestions that N400 effects occur too late to represent the first phase of lexical semantic access (van den Brink et al., 2001; Sereno and Rayner, 2003; Hauk et al., 2006; Dambacher et al., 2009).

CONCLUSIONS

In sum, initial access to semantic features associated with visual words is fast: ERP modulations set in already at about 190 ms, in close temporal succession to orthographic activation as indicated by lexicality effects in the N1, starting at about 164 ms. Furthermore, the amount of semantic features enhanced N400 amplitudes, indicating continued influences of word meaning during reading.

ACKNOWLEDGMENTS

This work was supported by a scholarship from the Berlin School of Mind and Brain to Milena Rabovsky, and the German Research Foundation Grants AB 277/4 and 5 to Rasha Abdel Rahman. We would like to thank Melih Bakirtas for assisting in data acquisition.

REFERENCES

Abdel Rahman, R., and Sommer, W. (2008). Seeing what we know and understand: how knowledge shapes perception. *Psychon. Bull. Rev.* 15, 1055–1063.

Adelman, J. S., Brown, G. D., and Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychol. Sci.* 17, 814–823.

Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia* 49, 970–983.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. Gen.* 133, 283–316.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behav. Res. Methods* 39, 445–459.

Barber, H. A., Ben-Zvi, S., Bentin, S., and Kutas, M. (2011). Parafoveal perception during sentence reading? An ERP paradigm using rapid serial visual presentation (RSVP) with flankers. *Psychophysiology* 48, 523–531.

Berg, P., and Scherg, M. (1991). Dipole modeling of eye activity and its application to the removal of eye artefacts from the EEG and MEG. *Clin. Phys. Physiol. Meas.* 12, 49–54.

Brandeis, D., Naylor, H., Halliday, R., Callaway, E., and Yano, L. (1992). Scopolamine effects on visual information processing, attention, and event-related potential map latencies. *Psychophysiology* 29, 315–336.

Brem, S., Halder, P., Bucher, K., Summers, P., Martin, E., and Brandeis, D. (2009). Tuning of the visual word processing system: distinct developmental ERP and fMRI effects. *Hum. Brain Mapp.* 30, 1833–1844.

Brunswick, N., McCrory, E., Price, C. J., Frith, C. D., and Frith, U. (1999). Explicit and implicit processing of words and pseudowords by adult developmental dyslexics. *Brain* 122, 1901–1917.

Buchanan, L., Westbury, C., and Burgess, C. (2001). Characterizing semantic space: neighborhood effects in word recognition. *Psychon. Bull. Rev.* 8, 531–544.

Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256.

Dambacher, M., Rolfs, M., Göllner, K., Kliegl, R., and Jacobs, A. M. (2009). Event-related potentials reveal rapid verification of predicted visual input. *PLoS One* 4, e5047. doi: 10.1371/journal.pone.0005047

Duñabeitia, J. A., Áviles, A., and Carreiras, M. (2008). NoA's ark: influence of the number of associates in visual word recognition. *Psychon. Bull. Rev.* 15, 1072–1077.

Fiez, J. A., Balota, D. A., Raichle, M. E., and Petersen, S. E. (1999). Effects of lexicality, frequency, and spelling-to-sound consistency of the functional anatomy of reading. *Neuron* 24, 205–218.

Grondin, R., Lupker, S. J., and McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *J. Mem. Lang.* 60, 1–19.

Harm, M., and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111, 662–720.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., and Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage* 30, 1383–1400.

Holcomb, P. J., Kounios, J., Anderson, J. E., and West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 721–742.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychon. Bull. Rev.* 10, 785–813.

Kandhadai, P., and Federmeier, K. D. (2010). Automatic and controlled aspects of lexical associative processing in the two cerebral hemispheres. *Psychophysiology* 47, 774–785.

Kiefer, M., and Pulvermüller, F. (2011). Conceptual representations in mind and brain: theoretical developments, current evidence, and future directions. *Cortex.* Advance online publication.

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., and Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *J. Neurosci.* 28, 12224–12230.

Kounios, J., Green, D. L., Payne, L., Fleck, J. I., Grondin, R., and McRae, K. (2009). Semantic richness and the activation of concepts in semantic memory: evidence from event-related potentials. *Brain Res.* 1282, 95–102.

Kounios, J., and Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 804–823.

Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470.

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 14.1–14.27.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.

Lehmann, D., and Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr. Clin. Neurophysiol.* 48, 609–621.

Locker, L. Jr., Simpson, G. B., and Yates, M. (2003). Semantic neighborhood effects on the recognition of ambiguous words. *Mem. Cognit.* 31, 505–515.

Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychon. Bull. Rev.* 7, 618–630.

Lund, K., and Burgess, C. (1996). Producing high-deimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* 28, 203–208.

Maurer, U., Brandeis, D., and McCandliss, B. D. (2005). Fast, visual specialization for reading in English revealed by the topography of the N170 ERP response. *Behav. Brain Funct.* 1, 13.

McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* 7, 293–299.

McRae, K. (2004). "Semantic memory: some insights from feature-based connectionist attractor networks," in *Psychology of Leaning and Motivation: Advances in Research and Theory,* Vol. 45, ed B. H. Ross (San Diego, CA: Academic Press), 41–86.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods Instrum. Comput.* 37, 547–559.

Mirman, D., and Magnusson, J. S. (2008). Attractor dynamics and semantic neighborhood density: processing is slowed by near neighbors and speeded by distant neighbors. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 65–79.

Müller, O., Duñabeitia, J. A., and Carreiras, M. (2010). Orthographic and associative neighborhood density effects: what is shared, what is different? *Psychophysiology* 47, 455–466.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behav. Res. Methods Instrum. Comput.* 36, 402–407.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.

Penolazzi, B., Hauk, O., and Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biol. Psychol.* 74, 374–388.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on visual word recognition. *Psychon. Bull. Rev.* 15, 161–167.

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., and Goodyear, B. G. (2007). The neural consequences of semantic richness. When more comes to mind, less activation is observed. *Psychol. Sci.* 18, 401–406.

Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics on visual word recognition: number-of-features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.

Pexman, P. M., Holyk, G. G., and MonFils, M. H. (2003). Number-of-features effects and semantic processing. *Mem. Cognit.* 31, 842–855.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115.

Plaut, D. C., and Shallice, T. (1993). Deep dyslexia – a case-study of connectionist neuropsychology. *Cogn. Neuropsychol.* 10, 377–500.

Pulvermüller, F., Shtyrov, Y., and Hauk, O. (2009). Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain Lang.* 110, 81–94.

Rabovsky, M., Sommer, W., and Abdel Rahman, R. (2011a). Depth of conceptual knowledge modulates visual processes during word reading. *J. Cogn. Neurosci.* Advance online publication.

Rabovsky, M., Sommer, W., and Abdel Rahman, R. (2011b). Implicit word learning benefits from semantic richness: electrophysiological and behavioral evidence. *J. Exp. Psychol. Learn. Mem. Cogn.* Advance online publication.

Schott, B., Richardson-Klavehn, A., Heinze, H. J., and Düzel, E. (2002). Perceptual priming versus explicit memory: dissociable neural correlates at encoding. *J. Cogn. Neurosci.* 14, 578–592.

Segalowitz, S., and Zheng, X. (2009). An ERP study of category priming: evidence of early lexical semantic access. *Biol. Psychol.* 80, 122–129.

Seidenberg, M. S., and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychol. Rev.* 96, 523–568.

Sereno, S. C., and Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends Cogn. Sci.* 7, 489–493.

Skrandies, W. (1998). Evoked potential correlates of semantic meaning – A brain mapping study. *Brain Res. Cogn. Brain Res.* 6, 173–183.

Swaab, T. Y., Baynes, K., and Knight, R. T. (2002). Separable effects of priming and imageability on word processing: an ERP study. *Brain Res. Cogn. Brain Res.* 15, 99–103.

van den Brink, D., Brown, C. M., and Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *J. Cogn. Neurosci.* 13, 967–985.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., and Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word form system. *Neuron* 55, 143–156.

West, W. C., and Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *J. Cogn. Neurosci.* 12, 1024–1037.

Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.

Yates, M., Locker, L. Jr., and Simpson, G. B. (2003). Semantic and phonological influences on the processing of words and pseudohomophones. *Mem. Cogn.* 31, 856–866.

Yee, E., Overton, E., and Thompson-Schill, S. L. (2009). Looking for meaning: eye movements are sensitive to overlapping features not association. *Psychon. Bull. Rev.* 16, 869–874.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 November 2011; accepted: 23 January 2012; published online: 14 February 2012.

Citation: Rabovsky M, Sommer W and Abdel Rahman R (2012) The time course of semantic richness effects in visual word recognition. Front. Hum. Neurosci. ***6****:11. doi: 10.3389/fnhum.2012.00011*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 03 April 2012
doi: 10.3389/fnhum.2012.00073

Richer concepts are better remembered: number of features effects in free recall

Ian S. Hargreaves*, Penny M. Pexman*, Jeremy C. Johnson and Lenka Zdrazilova

Department of Psychology, Language Processing Lab, University of Calgary, Calgary, AB, Canada

Edited by:
Paul D. Siakaluk, University of Northern British, USA

Reviewed by:
Michael E. J. Masson, University of Victoria, Canada
Carrie Esopenko, Rotman Research Institute, Canada

****Correspondence:***
Ian S. Hargreaves and Penny M. Pexman, Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada.
e-mail: ishargre@ucalgary.ca

Many models of memory build in a term for encoding variability, the observation that there can be variability in the richness or extensiveness of processing at encoding, and that this variability has consequences for retrieval. In four experiments, we tested the expectation that encoding variability could be driven by the properties of the to-be-remembered item. Specifically, that concepts associated with more semantic features would be better remembered than concepts associated with fewer semantic features. Using feature listing norms we selected sets of items for which people tend to list higher numbers of features (high NoF) and items for which people tend to list lower numbers of features (low NoF). Results showed more accurate free recall for high NoF concepts than for low NoF concepts in expected memory tasks (Experiments 1–3) and also in an unexpected memory task (Experiment 4). This effect was not the result of associative chaining between study items (Experiment 3), and can be attributed to the amount of item-specific processing that occurs at study (Experiment 4). These results provide evidence that stimulus-specific differences in processing at encoding have consequences for explicit memory retrieval.

Keywords: semantic richness, free recall, memory

Words vary on a large number of lexical dimensions that characterize factors such as their frequency of usage, or that refer to structural characteristics such as shape (orthography) and sound (phonology). Words, rather helpfully, also vary in meaning, and this variability can be captured by numerous semantic dimensions that influence the speed with which words can be recognized or categorized (Pexman et al., 2008). A vast word recognition literature has sought to characterize how orthographic, phonologic, and semantic dimensions interactively contribute to our ability to read. Consistently, researchers have shown that the variability of a given word along any or all of these dimensions is an important determinant in how that word is processed, manifesting in differences in reading times and accuracy (Yap and Balota, 2009). Words are also convenient stimuli for experiments, and are often utilized in memory research as they offer a well-defined minimal unit that can easily serve in recognition and free recall memory paradigms. This raises an interesting question: we know that there are many characteristics of individual words that shape how those words are processed, but do these item-specific differences influence subsequent memory when words are used as stimuli?

One approach to characterizing these effects is also one of the most influential frameworks in human memory research. The levels of processing framework proposed by Craik and Lockhart (1972) provided a number of important ideas, including the assertion that deeper processing at encoding leads to more accurate recollection at retrieval. In later work the framework was refined in a number of ways, and depth of processing was distinguished from another important type of encoding: elaboration. While depth of processing refers to the fact that some domains of processing typically involve richer or more extensive processing than others, elaboration has been characterized as "richness or extensiveness of processing within each qualitative type (of processing)" (Lockhart and Craik, 1990, p. 100). That is, within a particular type or domain of processing (e.g., semantic processing) there is variability in processing richness and this variability has consequences for memory. Numerous studies of semantic elaboration showed that free recall could be influenced by manipulating the encoding conditions applied to the to-be-remembered items (e.g., Craik and Tulving, 1975; Klein and Saltz, 1976; Ritchey and Beal, 1980; Ross, 1981; Hashtroudi, 1983) and importantly for the present discussion, by the variability in semantic elaboration prompted by the characteristics of the to-be-remembered items themselves (Seamon and Murray, 1976). This revised emphasis on elaboration helped to shift the levels of processing framework away from a focus on the depth of processing *per se* and toward a focus on how qualitatively distinct encoding operations influence memory. This shift was important, as the levels of processing framework was criticized for being underspecified (Morris et al., 1977) or worse, inherently circular (Nelson, 1977). However, despite this advancing construal of levels of processing, researchers continued to struggle with implementing the framework within a computational model (Eich, 1985; cf. Craik and Lockheart, 1986).

Researchers still show great interest in characterizing how variability in processing during encoding can influence subsequent memory. Indeed, the primary assertion of semantic elaboration (that the relative amount of processing within a single domain should predict subsequent memory) finds a more clearly specified counterpart in the construct of encoding variability[1]. Similar to elaboration, encoding variability captures the idea that variability in how items are processed will lead to differences in memory strength across items. This intuitive assumption has been implemented in models of recognition memory in order to account for the observation that studied items vary more in memory strength than new items (Hintzman, 1986; cf. Koen and Yonelinas, 2010). It has also been used to interpret the observation that brain-based changes at encoding predict the subsequent recall of items, for example item-wise variability in hippocampal gamma oscillations predict the likelihood of successful free recall (Sederberg et al., 2007). Encoding variability can also be implemented in models of free recall (Sederberg et al., 2008), offering a level of specification that the elaboration account lacks.

Both semantic elaboration and encoding variability literatures make the prediction that processing differences at encoding will lead to subsequent effects in free recall. However, neither has given much attention to potential differences in processing that are spontaneously elicited by the lexico-semantic characteristics of to-be-studied items. This is an important point; words are known to vary on a large number of lexico-semantic dimensions, and to the extent that this variability automatically shapes the processing of these items, both semantic elaboration and encoding variability accounts would predict subsequent effects in free recall.

In related research, Nelson and colleagues have investigated how the associative relationships between words can influence memory performance for individual words. In natural language usage words are produced in structured sentences that lead them to become entangled with one another. Nelson and colleagues captured these associative relationships by asking a large number of participants to list the first word that comes to mind in relation to a presented target word (Nelson et al., 1998). Using this database, Nelson and colleagues documented effects of words' Number of Associates (NoA; also known as associative set size) in a variety of memory tasks. Compared to words with many associates, words with fewer associates are more likely to be successfully retrieved during cued recall, however, manipulating NoA did not influence free recall performance (Nelson and Schreiber, 1992). That NoA influences cued but not free recall suggests that the influence of lexico-semantic variables on memory performance is likely task-specific. The concreteness variable shows a different pattern across tasks: relative to abstract words (e.g., *VIRTUE*), concrete words (e.g., *CAT*) show more accurate performance in cued recall, free recall and recognition memory tasks (Paivio and Csapo, 1973; Nelson and Schreiber, 1992; Hamilton and Rajaram, 2001). In visual word recognition, there have been repeated demonstrations that the effects of item-specific relative semantic richness are multidimensional, leading variables like NoA and concreteness to dissociate across different visual word recognition tasks (Pexman et al., 2008; Yap et al., 2011). While it is not surprising to observe similar dissociations in a task as unconstrained as free recall, the potential for lexico-semantic variables to selectively influence different memory tasks highlights the importance of properly balanced stimulus sets. Seamon and Murray (1976) manipulated subjectively rated meaningfulness, which uses a Likert-type scale to measure the extent to which participants feel that a word arouses other associated words (with more words leading to higher values; Toglia and Battig, 1978). Unfortunately, it is unclear what information participants use when placing words on a dimension of meaningfulness and this variable shows significant correlations with other subjectively rated variables such as familiarity, imageability, and concreteness. Indeed, in the Seamon and Murray study the high meaningful words were also high on ratings of imagery and concreteness. Because of the difficulty in operationalizing meaningfulness it is unknown whether this manipulation is fine-grained enough to test theories of elaboration, since high meaningful words may differ on any number of dimensions from low meaningful words. The goal of the current study was to investigate item-specific encoding variability in a more precise fashion than in previous studies, by investigating number-of-features (NoFs) effects in free recall.

NoF refers to the number of semantic features that participants list for different concepts in a feature-listing task (Pexman et al., 2002). The features listed for different concepts are considered "verbal proxies for packets of knowledge" (McRae, 2005, p. 42), rather than veridical descriptions of semantic memory. As they generate features, participants access representations derived from their experience with the concepts. McRae and colleagues (McRae et al., 2005) published feature norms for 541 concrete concepts. For instance, for the concept *cow* the normative features include perceptual characteristics like *has four legs, has an udder*, and *is smelly*. Other features describe behaviors, like *eats grass*, and *moos*. Some of the features describe the concept's function, like *produces milk,* or its context, as in *lives on farms*. There is variability in the number of features listed for different concepts (e.g., 20 for *couch*, 23 for *cougar*, 11 for *table*, 9 for *leopard*) and this variability is related to responding in word recognition tasks (lexical decision, semantic categorization), such that responses are faster and more accurate for words with many features than for words with few features, even when other variables, like word length, frequency, typicality, and concreteness, are controlled (Pexman et al., 2002, 2003, 2008; Grondin et al., 2009; Yap et al., 2011). The processing advantage observed for high NoF words has been attributed to greater semantic activation for high NoF concepts (Pexman et al., 2003).

NoF effects have only been examined in visual word recognition tasks. In the present study we investigated whether NoF effects can be observed in free recall. Compared to past investigations that manipulated meaningfulness and concreteness, the relative transparency with which the NoF variable is defined allowed us to test for fine-grained effects of item-specific encoding variability in memory performance. Given the nature of these effects as outlined above, one would expect that the enriched

[1]The term encoding variability also refers to a class of phenomena in the spacing effect literature in which encountering an item in numerous (or variable) contexts confers a memory advantage relative to items encountered in a single context (e.g., Waters and McAlaster, 1983). Here, we use the term solely to refer to variability in memory strength in the sense that some items are encoded very well, and this influences subsequent memory performance. Our items are balanced with respect to their normative distribution across textual contexts (Brysbaert and New, 2009).

encoding afforded by high NoF words would lead to more accurate recall. Of course, given the narrow definition of semantic richness captured by NoF, it was also possible that the difference between high and low NoF words would be too subtle to influence memory accuracy. To investigate these possibilities we chose free recall because an extensive literature shows that this task produces effects of another stimulus-specific property: concreteness (Dukes and Bastian, 1966; Paivio and Csapo, 1973; Nelson and Schreiber, 1992; Paivio et al., 1994; Ruiz-Vargas et al., 1996; Hamilton and Rajaram, 2001; ter Doest and Semin, 2005), and we modeled our procedure after the most recent of these studies. To be clear, however, we investigated NoF effects for sets of items for which concreteness, word frequency, familiarity, and contextual diversity was controlled, so any memory effects observed for NoF could be interpreted as incremental to those of each of these other factors. In Experiments 1 and 2 we tested for fine-grained effects of item-specific encoding variability by investigating whether NoF effects can be observed in free recall. In Experiments 3 and 4 we further explored the mechanisms for those effects by investigating whether NoF effects are the result of associative chaining among items rather than superior recall for individual items (Experiment 3) and by investigating whether NoF effects emerge during the incidental encoding of to-be-remembered items in a lexical decision task (LDT; Experiment 4).

EXPERIMENT 1

METHOD

Participants

Participants in Experiment 1 were 30 undergraduate students at the University of Calgary. In all of the experiments reported in this paper, participants reported that English was their first language, had normal or corrected-to-normal vision, and received course credit for participation.

Materials

The stimuli for Experiment 1 were 30 low NoF words and 30 high NoF words selected from the McRae et al. (2005) norms (**Table A1**). The selected word sets differed significantly in NoF ($p < 0.001$) but were matched for printed frequency, contextual diversity (Brysbaert and New, 2009), familiarity, printed length, orthographic neighborhood size (Coltheart et al., 1977), and concreteness (see **Table 1**). As a result of this matching, differences between the low NoF and high NoF words on each of these dimensions were non-significant at $p > 0.10$[2]. We obtained concreteness values for 55 of the items from the MRC database (Wilson, 1988), and collected concreteness ratings for the five remaining items from a separate group of 31 participants.

Procedure

There were three components in a testing session: (1) a study phase, (2) a distraction phase, and (3) a recall phase. On each trial in the study phase, a word was presented in the center of a 17" monitor controlled by a Macintosh G3 computer using PsyScope (Cohen et al., 1993). Each word was presented for 2 s, followed by 3 s of blank screen before presentation of the next word (ter Doest and Semin, 2005). A total of 60 words were presented for study, in a different random order to each participant. Participants were asked to memorize the words for a later recall test. In the distraction phase, participants were asked to complete two unrelated tasks on the computer: a semantic categorization task and a ratings task, both with word stimuli. The time taken to complete the

[2]In retrospect, we investigated possible issues with our stimulus sets. While the study item sets were matched on numerous lexical dimensions, subsequent examination of the High and Low NoF items revealed significant differences in the Number of Associates for the items used in Experiments 1 and 3, $t(58) = -2.69$, $p < 0.05$, SE = 1.82. In addition, we collected additional concreteness and new age of acquisition (AoA) ratings from separate groups of participants at the University of Calgary. AoA values were collected from a group of 144 undergraduate students. Each of these students provided AoA ratings for one-quarter of a larger set of 514 words, such that 36 students provided ratings for each word. The instructions for the AoA ratings task originated from Carroll and White (1973), but we used the modified 7-point scale of Gilhooly and Hay (1977). Concreteness ratings were collected for all 110 items used in Experiments 1 and 2 from a new set of 20 participants. For these new ratings data we detected small but statistically significant differences in concreteness and in AoA between the High and Low NoF items used in Experiments 1 and 3, $t(58) = 2.80$, $p < 0.05$, SE = 0.09, $t(58) = 2.02$, $p = 0.048$, SE = 0.27 respectively. NoA, concreteness and AoA were all balanced in the item set used in Experiments 2 and 4 (all non-significant $p > 0.10$). While studies have shown that NoA is an important determinant of cued recall, manipulating NoA does not influence free recall (Nelson and Schreiber, 1992). In addition, while the observed differences in rated concreteness and AoA between High and Low NoF items was significant, both sets of items are very concrete, and are perceived to be learned early on, between kindergarten and the first grade. This, and the observation that a NoF effect of similar size was observed across all four studies provides prima facie evidence that the NoF effect observed in Experiments 1 and 3 was not driven by these differences in AoA or concreteness.

Table 1 | Mean stimulus characteristics (standard deviations in parentheses).

Word type	Frequency	CD	Familiarity	Letter Length	Orth N	AoA	Conc (old)	Conc (new)	NoF	NoA
EXPERIMENTS 1 AND 3										
Low NoF	14.31 (24.24)	4.55 (6.39)	5.80 (1.96)	5.23 (1.52)	5.80 (6.29)	3.51 (1.24)	5.99 (0.28)	6.56 (0.43)	8.17 (1.12)	8.60 (7.35)
High NoF	18.96 (22.89)	5.81 (6.04)	5.76 (2.10)	5.30 (1.60)	5.43 (6.56)	2.95 (0.86)	6.07 (0.24)	6.82 (0.23)	17.47 (1.59)	13.53 (6.79)
EXPERIMENTS 2 AND 4										
Low NoF	10.34 (15.98)	3.37 (4.29)	6.49 (1.96)	5.96 (2.05)	3.32 (3.83)	3.07 (0.88)	6.24 (0.30)	6.88 (0.12)	10.96 (1.65)	10.40 (6.13)
High NoF	15.01 (16.12)	5.00 (4.82)	6.34 (2.11)	5.52 (1.58)	3.88 (4.23)	2.91 (0.99)	6.34 (0.25)	6.88 (0.16)	18.12 (1.72)	11.04 (7.39)

Note: CD, contextual diversity; NoF, number of features; orth N, orthographic neighborhood size; AoA, age of acquisition; Conc, concreteness; NoA, Number of Associates.

distraction tasks was 9 min. In the recall phase, participants were presented with a blank computer screen and were asked to try to remember the studied words, typing in each word they recalled. Participants were given 4 min to complete the recall phase but could request more time (Hamilton and Rajaram, 2001). None of the participants requested additional time.

Coding procedures for recall responses were adopted from those used in previous studies (e.g., ter Doest and Semin, 2005). Responses were judged correct if they were identical to, or were inflectional or misspelled variants of words on the study list (e.g., we accepted *shelf* for *shelves*, and *plyers* for *pliers*). Responses were judged incorrect if they did not appear on the study list or were synonyms of a studied word (e.g., we did not accept *cabinet* for *cupboard*).

RESULTS AND DISCUSSION

The mean proportions of low NoF and high NoF words recalled are presented in **Table 2**. In addition to the studied items, participants recalled an average of 2.80 words (SD = 3.08) that were not in the studied list. T-tests were conducted with subjects (t_1) and, separately, items (t_2) as random factors to compare correct recall for low and high NoF words. Results showed a significant NoF effect ($t_{1(29)} = 3.65$, $p < 0.001$, SE $= 0.02$; $t_{2(58)} = 2.91$, $p < 0.01$, SE $= 0.03$): recall was better for high NoF words than for low NoF words. This was, to our knowledge, the first report of a NoF effect in memory and we sought to replicate it with a different set of items in Experiment 2.

EXPERIMENT 2

METHOD

Participants

Participants in Experiment 2 were 37 undergraduate students at the University of Calgary.

Materials

The stimuli for Experiment 2 were the 25 high NOF words and 25 low NOF words used in Pexman et al. (2002) (**Table A2**). The selected word sets differed significantly in NoF ($p < 0.001$) but were matched for printed frequency, contextual diversity (Brysbaert and New, 2009), familiarity, printed length, orthographic neighborhood size (Coltheart et al., 1977), and concreteness (see **Table 1**). All matching was non-significant at $p > 0.10$. We obtained concreteness values for 26 of the present items from the MRC database (Wilson, 1988), and collected concreteness ratings for the 24 remaining items from a separate group of 31 participants. No participants requested additional time for free recall.

RESULTS AND DISCUSSION

The mean proportions of low NoF and high NoF words correctly recalled are presented in **Table 2**. In addition to the studied items, participants recalled an average of 2.49 words (SD = 3.08) that were not in the studied list. Results showed a significant NoF effect ($t_{1(36)} = 3.23$, $p < 0.005$, SE = 0.02; $t_{2(48)} = 2.01$, $p < 0.05$, SE = 0.04): the proportion of correctly recalled items was higher for high NoF words than for low NoF words. With the existence of the effect established and replicated, we next sought to investigate the source of the effect.

The free recall task has a long history in memory research (Kirkpatrick, 1894). Participants must engage in a selective search of memory in order to produce items studied at an earlier time. The unconstrained nature of this process means that multiple informational dimensions are free to interact with this search, yielding a long list of factors that influence recall dynamics. Factors such as the relative decay of items from memory (e.g., recency effects; Glanzer and Cunitz, 1966), additional rehearsal at study (e.g., primacy effects) and any factor that might influence the order with which information comes to mind, such as the order of presentation at study or semantic proximity to other items on the study list (e.g., contiguity effects; Kahana, 1996) all dynamically contribute to recall performance.

Recent work by Kahana and colleagues has produced models of immediate free recall that successfully incorporate many of these factors (Sederberg et al., 2008). Importantly, they also outline a mechanism for effects of item-specific encoding variability, and have the potential to account for the observation of a NoF effect in free recall. For example, the temporal context model (TCM-A) of Sederberg et al. (2008) frames free recall as the result of a series of stages. At study, the presentation of the to-be-studied items drives the evolution of a context layer which is stored in memory. Since item presentation drives the evolution of context, temporal information about the successive order of items, previous contexts associated with that item (i.e., semantic information), and information about the current context combine to create a context representation that can then guide later memory search. It is through this mechanism that the overall study context forms associations with the representations of the individual studied items, which enables subsequent retrieval of those items during recall. Free recall of items using these contextual states is modeled as a competitive process among a set of leaky accumulators (Usher and McClelland, 2001). Items that leave a stronger trace in the context layer at study will be more active during this process, and will be more likely to be produced during free recall. While TCM-A is a model of immediate recall, this framework provides a potential mechanism for effects

Table 2 | Mean proportion of words correctly recalled.

Word type	*M*	SD
EXPERIMENT 1		
Low NoF	0.25	0.16
High NoF	0.32	0.20
EXPERIMENT 2		
Low NoF	0.26	0.16
High NoF	0.33	0.16
EXPERIMENT 3		
Low NoF	0.27	0.16
High NoF	0.31	0.15
EXPERIMENT 4		
Low NoF	0.08	0.27
High NoF	0.15	0.35

Note: NoF, number of features.

of item-specific encoding variability such as those observed when manipulating NoF in delayed recall. By increasing item-specific activity at encoding we vary the relative ability of an item to bind itself to the prevailing context at study (Sederberg et al., 2008), and thus increase the likelihood of that item being active during subsequent recall.

If, following TCM-A, NoF effects in free recall result from variability in encoding at study, then we can make two predictions. First, the relative increase in an item's ability to bind itself to context should be specific to that item. Any subsequent benefit in free recall for that item should be driven by its improved ability to compete during free recall, not by any form of contiguity effect in which the temporal or semantic relationships between items at study influence retrieval, thereby creating associative chains between items that are recapitulated in free recall (Polyn et al., 2011). Thus, while we should very likely observe contiguity effects in recall, these contiguity effects should not be stronger for high NoF words than for low NoF words. Second, the locus of the NoF effect should be at encoding, and thus the quantifiable amount of semantic processing for an item that occurs at study should predict the likelihood of that item being recalled. These predictions were investigated in Experiment 3 and Experiment 4.

EXPERIMENT 3

In Experiment 3 we investigated whether NoF effects arise due to associative chaining between studied items as a function of NoF. As such we recorded the sequential ordering of item presentation at study (something we did not do in Experiments 1 or 2). It is worthwhile to note that Experiment 3 was not designed as a strong test of contiguity effects in free recall; there is substantial evidence that such effects are genuine (Kahana, 1996; Polyn et al., 2011). Rather, Experiment 3 was designed to test whether the NoF effect observed in Experiment 1 was the result of associative chaining between items due to contiguity, or whether it resulted from enhanced item-specific encoding.

METHOD

Participants

Participants in Experiment 3 were 42 undergraduate students at the University of Calgary.

Materials

The stimuli used in Experiment 3 were the same 60 items used in Experiment 1.

Procedure

The procedure for Experiment 3 was largely the same as that described for Experiment 1, but here we used a single distracter task during the distracter phase. In the distracter phase participants made semantic categorization decisions to single words presented on the monitor, again for 9 min. Stimuli were presented using *E*-Prime presentation software (Psychological Software Tools, Pittsburgh, PA) on a 19 inch dell monitor. We used the same coding procedures outlined in Experiment 1. No participants requested additional time.

RESULTS AND DISCUSSION

The mean proportions of low NoF and high NoF words recalled are presented in **Table 2**. In addition to the studied items, participants recalled an average of 4.14 words (SD = 3.41) that were not in the studied list. Results showed a NoF effect that was significant by subjects but not by items ($t_{1(42)} = 2.09$, $p < 0.05$, SE = 0.01; $t_{2(58)} = 0.27$, $p = 0.78$, SE = 0.02): recall was again better for high NoF words than for low NoF words.

Following Kahana (1996) and Ozubko and Joordens (2007), we constructed conditional response probability plots in order to reveal any association by contiguity. We plotted the probability of recalling an item that was between one to five positions ahead of or behind the just recalled item. A within-subjects ANOVA using these positional conditional probability plots revealed a significant effect of position using Greenhouse–Geisser corrections ($F_{(4.41,\ 176.60)} = 4.73$, $p = 0.001$, MSE = 0.01), this significant effect of contiguity indicates that participant recall was influenced by the sequential ordering of items at study. As **Figure 1** demonstrates, there was a large probability that a just-recalled item was one study position ahead of a previously recalled item. This pattern, along with a general bias to recall items in the forward direction, is typical of contiguity effects in free recall (Kahana, 1996; Ozubko and Joordens, 2007).

Since overall participant recall was influenced by contiguity we next turned to the question of whether the observed NoF effect was also a result of associative mechanisms that operate across items. Following Ozubko and Joordens (2007), for each participant we calculated the conditional probabilities of recalling a high NoF or low NoF item next, given that participants had just recalled a high or low NoF item. Averaging across participants yielded four conditional probabilities (presented in **Table 3**) that are sensitive to associative chaining among items as a function of NoF; *P*(highNoF|highNoF), *P*(lowNoF|highNoF), *P*(highNoF|lowNoF), and *P*(lowNoF|lowNoF). Using this notation, *P*(highNoF|lowNoF) would reflect the probability of recalling a high NoF item, having just recalled a low NoF item. If associations among items only form as a result of temporal

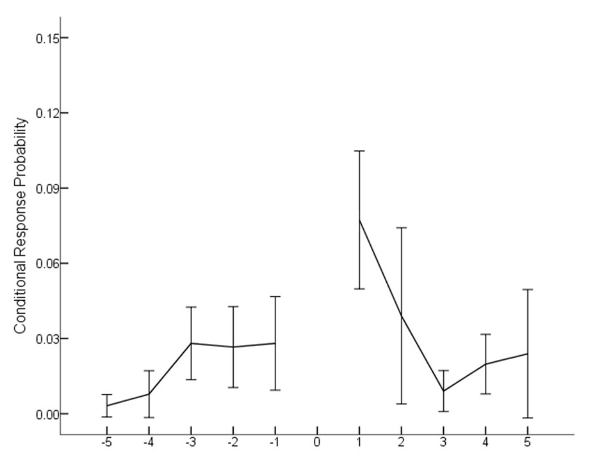


FIGURE 1 | Mean conditional probability plotted against distance from last item recalled.

Table 3 | Mean probability of recalling a High vs. a Low NoF word next, given that a High or Low NoF word has just been recalled (standard deviations in parenthesis).

Recalled	High next	Low next
	M	***M***
High	0.38 (0.19)	0.32 (0.18)
Low	0.32 (0.19)	0.33 (0.19)

Note: NoF, number of features.

proximity at study (i.e., the demonstrated contiguity effect) then these conditional probabilities defined in reference to NoF should be approximately equivalent. However, as demonstrated with word frequency effects (Ozubko and Joordens, 2007), varying NoF may lead to associative chaining among items. In this case, having just recalled a high or low NoF item would alter the probability of recalling either a high or low NoF item next, and significant differences among the four conditional probabilities should be observed. In order to account for the observation of a high NoF advantage in free recall, associative chaining as a function of NoF would have to take one of two forms. The first would be associative chaining among high NoF items, leading to an increased probability of recalling a high NoF item when having just recalled a high NoF item. In this situation, the *P*(highNoF|highNoF) should be significantly greater than the *P*(lowNoF|highNoF). Alternatively, the high NoF advantage could be explained via a decrease in associative chaining among low NoF items, thereby increasing the overall likelihood of producing high NoF words during free recall. In this situation, the *P*(lowNoF|lowNoF) should be significantly less than *P*(highNoF|lowNoF). To provide a test for this associative chaining as a function of NoF we used a paired-samples *t*-test to contrast the conditional probabilities *P*(highNoF|highNoF), *P*(lowNoF|highNoF), *P*(highNoF|lowNoF), and *P*(lowNoF|lowNoF) listed in **Table 3**. The results revealed that the conditional probabilities did not differ as a function of NoF, specifically, there was no evidence for differential associative chaining among high NoF items: having just recalled a high NoF word, participants were just as likely to recall a high NoF word (38%) as they were a low NoF word [32%; $t_{(41)} = 1.33$, $p = 0.19$, SE = 0.05]. Similarly, there was no evidence for reduced associative chaining among low NoF items: having just recalled a low NoF word, participants were just as likely to recall a low NoF word (33%) as they were to recall a high NoF word (32%; $t_{(41)} = 0.31$, $p = 0.76$, SE = 0.05). Indeed, plots of associative chaining by contiguity for both high and low NoF items (**Figure 2**) resemble the plots for the overall data (**Figure 1**). A within-subjects ANOVA using within-NoF positional conditional probability plots revealed a significant effect of position using Greenhouse–Geisser corrections for high NoF items, $F_{(3.95,\ 142.27)} = 2.77$, $p = 0.03$, MSE = 0.03. The same analysis for low NoF items (which are fewer in number, since fewer low NoF items were correctly recalled) revealed marginally significant results, $F_{(2.46,\ 83.63)} = 2.61$, $p = 0.06$, MSE = 0.03. Clearly, both high and low NoF items are capable of showing some degree of association by contiguity, including the classic asymmetrical bias

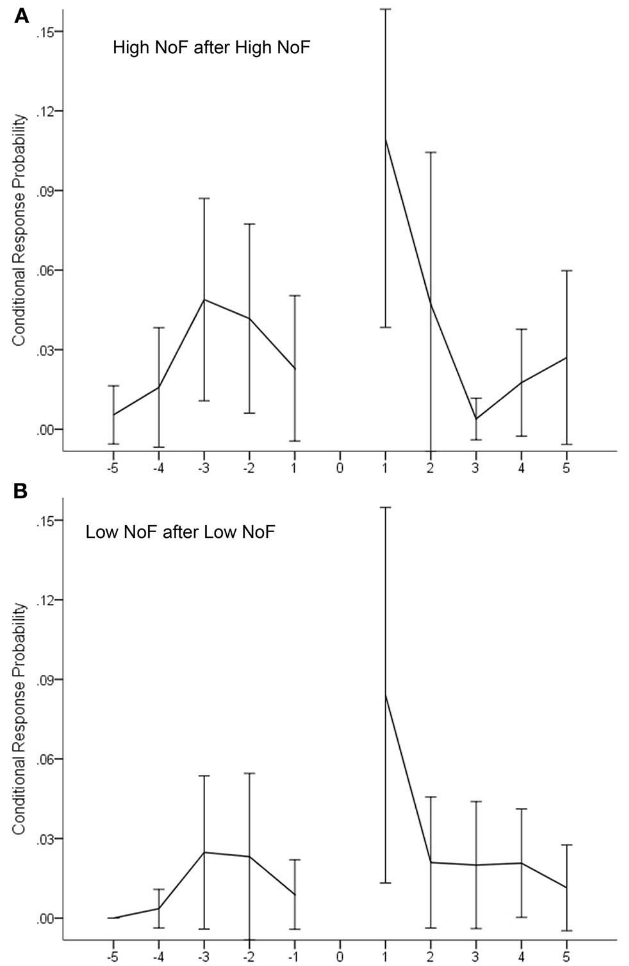


FIGURE 2 | Mean conditional probability plotted against distance from last item recalled for (A) High|High NoF and (B) Low|Low NoF items.

in favor of recalling items from study list positions that are nearer to the just-recalled-item.

Given that we observed no significant evidence for associative chaining as a function of NoF, one could argue that our tests simply lacked power. We conducted a *post-hoc* power analysis based on the effect size reported for the associative chaining in the low-frequency advantage in free recall reported by Ozubko and Joordens (2007). Like NoF, word frequency is another stimulus-specific variable that has been shown to influence free recall. These calculations suggested that only 32 participants would be required in order for our paired-sample comparisons to reach statistical significance. Since we tested 42 participants this indicates that our contrasts were sensitive enough to detect association by contiguity that varied as a function of NoF. Again, it is important to note that our goal for these analyses was to explore whether differential association by contiguity provides an explanation for the observation of a NoF advantage in Experiments 1, 2, and 3, where items were presented randomly, and later freely recalled. Under these specific conditions, the bulk of the evidence suggests that differential association by contiguity does not account for the NoF effect.

These results suggest that NoF effects in free recall do not arise from the associative chaining of high or low NoF items. Rather, the lack of differential associative contiguity among items as a function of NoF provides evidence that NoF effects arise from item-specific encoding variability. A stronger test of this conclusion would require a demonstration that the extensiveness of item-specific processing at encoding predicts the likelihood of recall. Experiment 4 was designed to test this prediction. In Experiment 4 we also investigated whether the NoF advantage in recall generalizes beyond the intentional learning paradigm used in the three experiments reported thus far. On the one hand, the NoF advantage may arise because participants are able to engage in more elaborative encoding for high NoF words during intentional learning of those items in the study phase. On the other hand, the NoF advantage may arise due to more extensive activation of the semantic system that occurs when high NoF words are processed. In the former case, NoF effects should arise only in an expected memory test (intentional memory). In the latter case, NoF effects should also arise in an unexpected memory test (incidental memory). This possibility was tested in Experiment 4.

EXPERIMENT 4

The goal of Experiment 4 was to investigate NoF effects in unexpected recall. In an initial version of this experiment we copied the procedure of Experiment 1, but changed the study phase task to a lexical decision task (LDT) and did not tell participants that they would need to recall the LDT word items later. The distraction tasks and timing were the same as in Experiment 1; that is, a 9 min distraction phase involving word judgment tasks. Results for this version of the experiment showed very poor recall performance (<3% items correctly recalled) and high rates of intrusion (participants recalled many items from the distraction tasks). To make the experiment somewhat easier and to reduce intrusions, in Experiment 4 we used a shorter distraction task comprised of math problems.

METHOD

Participants

Participants in Experiment 4 were 32 undergraduate students at the University of Calgary.

Materials

The stimuli for Experiment 4 were the same as in Experiment 2. There were also 50 non-words in the LDT.

Procedure

There were three components in a testing session: (1) a LDT, (2) a distraction phase, and (3) a recall phase. For the LDT participants first completed eight practice trials. Participants were told to decide whether each letter string presented in this task was a real word or a non-word, and to make their decision as quickly and as accurately as possible. Participants were *not* told that they would need to remember the LDT stimuli for a later phase of the session. In the distraction phase, participants completed a set of math problems. The time taken to complete the distraction task was 6 min. In the recall phase, participants were asked to try to remember as many of the LDT words as possible. Participants were given 4 min to complete the recall phase but could request more time. No participants requested additional time for free recall.

RESULTS AND DISCUSSION

LDT

LDT responses that were incorrect (3.2% of trials), or that were faster than 250 ms or slower than 2500 ms (less than 1% of trials) were excluded from the RT analysis. Mean RTs and errors are presented in **Table 4**. Results included a significant NoF effect in RT ($t_{1(31)} = 4.14$, $p < 0.001$, SE = 12.55; $t_{2(48)} = 2.16$, $p < 0.05$, SE = 29.14) but not in errors (both $t < 1$). This was the typical NoF effect in LDT; responses were faster for high NoF words than for low NoF words.

Recall

The mean proportions of low NoF and high NoF words recalled are presented in **Table 2**. In addition to the studied items, participants recalled an average of 1.72 words (SD = 1.68) that were not in the studied list. Results showed a significant NoF effect ($t_{1(31)} = 3.88$, $p < 0.001$, SE = 0.02; $t_{2(48)} = 2.72$, $p < 0.01$, SE = 0.02): recall was again better for high NoF words than for low NoF words.

NoF effects in visual word recognition are thought to capture the contributions of semantic processing to performance

Table 4 | Means and intercorrelations for LDT and recall performance in Experiment 4.

	M	SD	1	2	3	4	5	6	7	8
1. LDT RT (ms)—low NoF	677	287	–							
2. LDT RT (ms)—high NoF	627	225	0.89**	–						
3. LDT RT (ms)—NoF effect	50	71	0.64**	0.21	–					
4. LDT error—low NoF	0.03	0.17	−0.13	−0.24	0.14	–				
5. LDT error—high NoF	0.03	0.17	−0.21	−0.29	0.04	0.80**	–			
6. LDT error—NoF effect	0.00	0.03	0.16	0.13	0.13	0.13	−0.50**	–		
7. Recall accuracy—low NoF	0.08	0.27	0.14	−0.05	0.37*	−0.09	0.02	−0.15	–	
8. Recall accuracy—high NoF	0.15	0.35	0.05	−0.18	0.41*	0.11	−0.01	0.18	0.27	–
9. Recall accuracy—NoF effect	0.07	0.09	−0.04	−0.14	0.15	0.16	−0.02	0.27	−0.38*	0.79**

**$p < 0.05$; ** $p < 0.01$.

(Pexman et al., 2002). Therefore, we also examined the relationships between LDT performance and recall performance (**Table 4**) on the assumption that the magnitude of the NoF effect in LDT provides an index of the extensiveness of semantic processing at study. This analysis is designed to test whether the magnitude of the NoF effect shown by an individual participant predicted the subsequent recall of items. Notably, larger NoF effects in LDT RTs were related to better recall for both low NoF words and high NoF words. It is worth commenting on the non-significant relationship between the size of the NoF effect in LDT RTs and the size NoF effect in free recall. Across four Experiments, while we consistently found a NoF effect in free recall, the size of the NoF effect was consistently small (between 4% and 7%). This reduction in variability likely limits our ability to detect any significant correlation between RT and the magnitude of the NoF effect in free recall. While there is no evidence that the extent of semantic encoding during the LDT study phase is directly related to the *size* of the NoF effect in free recall, there is evidence linking the extent of semantic processing at study to recall accuracy for both low and high NoF words. This is a critical point, as it suggests that variability in the extensiveness of semantic processing undertaken by our participants during study is related to how much information they will subsequently recall. Given the careful balancing of the items in Experiment 4, we can reasonably attribute this variability in processing during encoding to the relative stimulus specific differences in NoF. While strictly correlative, this provides a tenable explanation for the NoF effect in free recall that is consistent with the literature on encoding variability. That the extensiveness of semantic processing was related to recall performance supports the inference that more extensive semantic processing at encoding leads to more accurate retrieval.

GENERAL DISCUSSION

The purpose of the present study was to investigate whether memory accuracy was modulated by stimulus-specific differences in encoding variability. We chose to manipulate the number of features (NoF) in order to elicit a shift in the relative processing among items at encoding. Results of four experiments showed that recall was more accurate for high NoF words than for low NoF words. An investigation into the mechanism of these NoF effects suggested that the observed benefit was due to stimulus-specific differences in encoding at study and did not result from associative chaining by NoF word type across studied items. Further, correlational results revealed that memory accuracy was related to the extent of semantic processing undertaken in the encoding task (as captured by the NoF effect in LDT), but was not related to the time spent processing the items at encoding. These results serve to constrain alternative hypotheses about the locus of NoF effects in free recall, and provide additional evidence that NoF effects are effects of item-specific encoding variability. Prior to this study the NoF effect had only been observed in word recognition tasks. The fact that the NoF effect generalizes to memory tasks suggests that the NoF dimension captures substantial variability in semantic processing.

We believe that the observed NoF effects in free recall provide a novel demonstration of semantic elaboration as proposed by the levels of processing framework, but as a framework, levels of processing does not actually provide a mechanism to account for these effects. Recent computational models of free recall such as TCM-A can, however, be modified to provide a mechanism for NoF effects by including a term that captures encoding variability at study. Across numerous tasks in which participants read individual words, the relative NoF of an item has been shown to influence the relative lexical processing of that item (Pexman et al., 2008; Yap et al., 2011). By modulating item-specific activity at encoding, NoF may vary the relative ability of an item to bind itself to the prevailing context at study (Sederberg et al., 2008). In TCM-A, free recall is modeled as a competitive process between items (Usher and McClelland, 2001) therefore, items that leave a stronger trace in memory will be more active and more likely to be produced during free recall. It is through this mechanism that variability in NoF would drive the encoding variability that ultimately leads to the observation of a NoF effect in free recall. In considering ingredients for a more complete model of free recall, Sederberg et al. (2008) outline such a mechanism, suggesting that if one assumes that items vary in the weighting of newly learned experimental item-to-context associations, one can model effects of encoding variability such as elaboration. The results of the current study provide behavioral evidence that the inclusion of a mechanism to account for encoding variability makes an important contribution to a models' characterization of free recall performance.

As reviewed earlier, the unconstrained nature of free recall allows for a number of factors to contribute to memory performance. In the current study, we examined the hypothesis that NoF effects are a form of item-specific encoding variability by examining whether there was a relationship between semantic processing at study and subsequent recall, and whether manipulating NoF lead to the formation of associations across items as a function of their NoF status. The balance of the evidence suggests that NoF effects are based in elaboration, not association. However, while we controlled for a number of lexico-semantic factors, the correlational structure of the English language virtually guarantees that our manipulation also encompasses some other undefined semantic relationship. For the present purpose, our interest was only whether a manipulation of NoF was sufficiently fine-grained to elicit a shift in the relative encoding of a set of to-be-remembered items. The results of Experiment 4 demonstrate that the amount of semantic processing at study directly predicts accuracy in free recall, and thus provide evidence in favor of this hypothesis, but alternative explanations are also possible. Howard and Kahana (2002) demonstrated that semantic similarity (calculated from LSA; Landauer and Dumais, 1997) led to significant semantic clustering in free recall, with items that had similar LSA vectors more likely to show association in free recall. In future research it would be useful to investigate whether other dimensions of semantic richness which have been shown to influence word recognition performance, such as number of semantic neighbors and contextual dispersion (see Pexman et al., 2008, for a review), are also related to memory task performance.

Gallo et al. (2008) argued that the degree of richness or elaboration achieved within a given level of processing will have consequences for memory performance because of distinctiveness. That is, they argued that when more features are encoded for a given stimulus, memory for that stimulus is more distinctive. Gallo et al. did not directly test the effects of "more features" on memory performance but we did so here. Our results confirm that, even when items are equated in all other ways, items that activate more semantic features are better remembered.

REFERENCES

Brysbaert, M., and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990.

Carroll, J. B., and White, M. N. (1973). Age-of-acquisition norms for 220 picturable nouns. *J. Verb. Learn. Verb. Behav.* 12, 563–576.

Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behav. Res. Methods* 25, 257–271.

Coltheart, M., Davelaar, E., Jonasson, J. T., and Besner, D. (1977). "Access to the internal lexicon," in *Attention and Performance VI*, ed S. Dornic (Hillsdale, NJ: Lawrence Erlbaum Associates), 535–555.

Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verb. Learn. Verb. Behav.* 11, 671–684.

Craik, F. I. M., and Lockheart, R. S. (1986). CHARM is not enough: comments on Eich's model of cued recall. *Psychol. Rev.* 93, 360–364.

Craik, F. I. M., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* 104, 268–294.

Cree, G. S., McNorgan, C., and McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: implications for theories of semantic memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 643–658.

Dukes, W. F., and Bastian, J. (1966). Recall of abstract and concrete words equated for meaningfulness. *J. Verb. Learn. Verb. Behav.* 5, 455–458.

Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychol. Rev.* 92, 1–38.

Gallo, D. A., Meadow, N. G., Johnson, E. L., and Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: evidence from the criterial recollection task. *J. Mem. Lang.* 58, 1095–1111.

Gilhooly, K. J., and Hay, D. (1977). Imagery, concreteness, age-of-acquisition, familiarity, and meaningfulness values for 205 five-letter words having single-solution anagrams. *Behav. Res. Methods Instrum.* 9, 12–17.

Glanzer, M., and Cunitz, A. (1966). Two storage mechanisms in free recall. *J. Verb. Learn. Verb. Behav.* 5, 351–360.

Grondin, R., Lupker, S. J., and McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *J. Mem. Lang.* 60, 1–19.

Hamilton, M., and Rajaram, S. (2001). The concreteness effect in implicit and explicit memory tests. *J. Mem. Lang.* 44, 96–117.

Hashtroudi, S. (1983). Type of semantic elaboration and recall. *Mem. Cognit.* 11, 476–484.

Hintzman, D. L. (1986). "Schema Abstraction" in a multiple-trace memory model. *Psychol. Rev.* 93, 411–428.

Howard, M. W., and Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *J. Mem. Lang.* 46, 85–98.

Kahana, M. (1996). Associative retrieval processes in free recall. *Mem. Cognit.* 24, 103–109.

Kirkpatrick, E. A. (1894). An experimental study of memory. *Psychol. Rev.* 1, 602–609.

Klein, K., and Saltz, E. (1976). Specifying the mechanisms in a levels-of-processing approach to memory. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 671–679.

Koen, J. D., and Yonelinas, A. P. (2010). Memory variability is due to the contribution of recollection and familiarity, not to encoding variability. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1536–1542.

Landauer, T. K., and Dumais, S. T. (1997). Solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240.

Lockhart, R. S., and Craik, F. I. M. (1990). Levels of processing: a retrospective commentary on a framework for memory research. *Can. J. Psychol.* 44, 87–112.

McRae, K. (2005). "Semantic memory: some insights from feature-based connectionist attractor networks," in *The Psychology of Learning and Motivation: Advances in Research and Theory, 45,* ed B. H. Ross (San Diego, CA: Elsevier Academic Press), 41–86.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods* 37, 547–559.

Morris, C. D., Bransford, J. D., and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *J. Verb. Learn. Verb. Behav.* 16, 519–533.

Nelson, T. O. (1977). Repetition and depth of processing. *J. Verb. Learn. Verb. Behav.* 16, 151–171.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). *The University of South Florida Word Association, Rhyme, and Word Fragment Norms.* Retrieved 2 Nov 2005, from http://www.usf.edu/FreeAssociation

Nelson, D. L., and Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *J. Mem. Lang.* 31, 237–260.

Ozubko, J., and Joordens, S. (2007). The mixed truth about frequency effects on free recall: effects of study list composition. *Psychon. Bull. Rev.* 15, 871–876.

Paivio, A., and Csapo, K. (1973). Picture superiority in free recall: imagery or dual coding? *Cogn. Psychol.* 5, 176–206.

Paivio, A., Walsh, M., and Bons, T. (1994). Concreteness effects on memory: when and why? *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 1196–1204.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on visual word recognition. *Psychon. Bull. Rev.* 15, 161–167.

Pexman, P. M., Holyk, G. G., and Monfils, M.-H. (2003). Number of features effects and semantic processing. *Mem. Cognit.* 31, 842–855.

Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number of features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.

Polyn, S., Erlikhman, G., and Kahana, M. (2011). Semantic cuing and the scale insensitivity of recency and contiguity. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 766–775.

Ritchey, G. H., and Beal, C. R. (1980). Image detail and recall: evidence for within-item elaboration. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 66–76.

Ross, B. H. (1981). The more, the better? Number of decisions as a determinant of memorability. *Mem. Cognit.* 9, 23–33.

Ruiz-Vargas, J. M., Cuevas, I., and Marschark, M. (1996). The effects of concreteness on memory: dual codes or dual processing? *Eur. J. Cogn. Psychol.* 8, 45–72.

Sederberg, P., Howard, M., and Kahana, M. (2008). A context-based theory of recency and contiguity in free recall. *Psychol. Rev.* 115, 893–912.

Sederberg, P., Shulze-Bonhage, A., Madsen, J., Bromfield, E., Litt, B., Brandt, A., and Kahana, M. (2007). Gamma oscillations distinguish true from false memory. *Psychol. Sci.* 18, 927–932.

Schwanenflugel, P. J., Akin, C., and Luh, W. (1992). Context availability and the recall of abstract and concrete words. *Mem. Cognit.* 20, 96–104.

Seamon, J. G., and Murray, P. (1976). Depth of processing in recall and recognition memory: differential effects of stimulus meaningfulness and serial position. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 680–687.

Toglia, M. P., and Battig, W. F. (1978). *Handbook of Semantic Word Norms*. Hillsdale, NJ: Erlbaum.

ter Doest, L. T., and Semin, G. R. (2005). Retrieval contexts and the concreteness effect: dissociations in memory for concrete and abstract words. *Eur. J. Cogn. Psychol.* 17, 859–881.

Usher, M., and McClelland, J. (2001). On the time course of perceptual choice: the leaky competing accumulator model. *Psychol. Rev.* 108, 550–592.

Waters, H., and McAlaster, R. (1983). Encoding variability and organization in free recall. *J. Exp. Child Psychol.* 36, 380–395.

Wilson, M. D. (1988). The MRC psycholinguistic database: machine readable dictionary, version 2. *Behav. Res. Methods Instrum. Comput.* 20, 6–11.

Yap, M. J., and Balota, D. A. (2009). Visual word recognition of multisyllabic words. *J. Mem. Lang.* 60, 502–529.

Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 January 2012; accepted: 16 March 2012; published online: 03 April 2012.

Citation: Hargreaves IS, Pexman PM, Johnson JC and Zdrazilova L (2012) Richer concepts are better remembered: number of features effects in free recall. Front. Hum. Neurosci. ***6****:73. doi: 10.3389/fnhum.2012.00073*

APPENDIX

Table A1 | Experiments 1 and 3 stimuli.

Low NoF words	High NoF words
Ball	Balloon
Bedroom	Barn
Birch	Basement
Biscuit	Bathtub
Bucket	Bear
Cabinet	Bus
Catfish	Cannon
Clamp	Canoe
Cod	Carrot
Doll	Cougar
Dove	Cow
Inn	Desk
Mackerel	Drapes
Mixer	Fawn
Otter	Gorilla
Parka	Grenade
Pine	Hammer
Pot	Kettle
Razor	Necklace
Rhubarb	Nylons
Rock	Pearl
Rocker	Pen
Scissors	Pickle
Shawl	Pig
Taxi	Rat
Toilet	Seal
Toy	Swimsuit
Trolley	Sword
Turnip	Toad
Veil	Train

Table A2 | Experiments 2 and 4 stimuli.

Low NoF words	High NoF words
Airplane	Apple
Broccoli	Bike
Catapult	Boots
Cherry	Bra
Closet	Cat
Corn	Coconut
Crayon	Couch
Crow	Dolphin
Cupboard	Fawn
Curtains	Freezer
Dresser	Fridge
Hawk	Garlic
Leotards	Goat
Lime	Grapefruit
Pillow	Lion
Pliers	Mouse
Pumpkin	Ostrich
Sandpaper	Pants
Scooter	Pistol
Shelves	Potato
Slippers	Screws
Stone	Sheep
Stove	Spoon
Truck	Tiger
Yam	Trousers

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 31 January 2013
doi: 10.3389/fnhum.2013.00005

ERP measures of semantic richness: the case of multiple senses

Vanessa Taler[1,2]*, Shanna Kousaie[2] and Rocío López Zunini[1,2]

[1] *School of Psychology, University of Ottawa, ON, Canada*
[2] *Bruyère Research Institute, Ottawa, ON, Canada*

Edited by:
Melvin Yap, National University of Singapore, Singapore

Reviewed by:
Milena Rabovsky, Humboldt-Universität zu Berlin, Germany
Ian S. Hargreaves, University of Calgary, Canada

****Correspondence:***
Vanessa Taler, School of Psychology, University of Ottawa, 136 Jean Jacques Lussier, Vanier Hall, Ottawa, ON K1N 6N5, Canada.
e-mail: vtaler@uottawa.ca

Semantic richness refers to the amount of semantic information that a lexical item possesses. An important measure of semantic richness is the number of related senses that a word has (e.g., TABLE meaning a piece of furniture, a table of contents, to lay aside for future discussion, etc.). We measured electrophysiological response to lexical items with many and few related senses in monolingual English-speaking young adults. Participants performed lexical decision on each item. Overall, high-sense words elicited shorter response latencies and smaller N400 amplitudes than low-sense words. These results constitute further evidence of the importance of semantic richness in lexical processing, and provide evidence that processing of multiple related senses begins as early as 200 milliseconds after stimulus onset.

Keywords: semantic richness, event-related potentials, N400, metonymy, lexical ambiguity

INTRODUCTION

When a reader recognizes a written word, information about the word's meaning is activated. Words differ in the amount of meaning-related (semantic) information that they possess, or their *semantic richness*. These differences can be measured in various ways. The first is the number of features generated in feature-listing tasks, where participants are asked to list all the semantic features associated with a given word (e.g., for the item KNIFE, these might include "a utensil," "is sharp," "is found in kitchens"). A second measure is the number of associates provided in free association tasks, where participants provide lexical associates for a given word (e.g., BIRD—NEST, BIRD—CAT). A third measure, central in understanding semantic networks, is the number of related meanings—or *senses*—a word possesses. For example, the word TABLE has many related senses (a piece of furniture, a table of contents, to lay aside for future discussion, etc.) while the word GUITAR refers only to a musical instrument.

A growing body of literature indicates that differences in semantic richness lead to different patterns of activation during word recognition. Measures of semantic richness predict response latency in lexical decision, naming, semantic categorization, and self-paced reading tasks (Pexman et al., 2002, 2003, 2008). A functional magnetic resonance imaging (fMRI) study found that semantically rich lexical items elicit less neural activation in left inferior frontal and temporal gyri than words that are more semantically impoverished (Pexman et al., 2007). In combination, these results indicate that increased semantic richness results in more rapid and less effortful lexical processing.

A critical question with respect to semantic richness effects is the timecourse of processing of this information. Event-related potentials (ERPs) are an ideal methodology to investigate this question. ERPs provide a real-time measure of neural processing, with millisecond-level resolution. Several recent studies have used ERPs to elucidate our understanding of the timecourse of activation of information related to semantic richness, including number of features and number of lexical associates (Kounios et al., 2009; Müller et al., 2010; Amsel, 2011; Laszlo and Federmeier, 2011; Rabovsky et al., 2012). In general, the component of interest has been the N400, a negative-going component in the ERP waveform that is traditionally assumed to reflect semantic processing, and whose negativity is inversely related to the semantic expectancy of a word (Kutas and van Petten, 1994). However, the N400 has also been found to be sensitive to context-independent semantic factors (Kounios, 1996) such as concreteness (Kounios and Holcomb, 1994). Semantic priming paradigms have demonstrated that the facilitated processing of a target word that is preceded by a related prime results in a smaller amplitude N400, suggesting that N400 amplitude is negatively related to the ease of processing a word.

With respect to semantic richness, a number of studies have found effects of various types of richness on the ERP waveform. Kounios et al. (2009) found marginally larger amplitude N400s to low- than high-number of feature words; the P2 component, peaking at around 240 milliseconds post-stimulus onset, was also influenced by number of features, with greater amplitude for high- than low-feature words. Rabovsky et al. (2012) investigated two subtypes of semantic richness: number of features and number of associates. Participants performed a lexical decision task on lexical items varying in these two measures. Number of associates did not exert an effect on the ERP waveform, while effects of number of semantic features were observed at central electrodes from 190 to around 500 milliseconds. Like Kounios et al., Rabovsky et al. observed a larger positive peak to low- than high-feature items in the P2/N2 window, arising at around 190 milliseconds post-stimulus onset. High-feature items, however, elicited a larger N400 than low-feature items—in contrast to the findings of Kounios et al. The authors conclude that initial semantic access occurs early and continues to exert an effect on processing at later stages in the reading process, and that partial data transmission occurs between the orthographic and semantic levels prior to completion of orthographic processing (i.e., processing is interactive rather than modular).

A more in-depth examination of semantic feature effects was reported by Amsel (2011), who compared the effect of different semantic feature types on the ERP waveform. Effects were observed prior to 200 milliseconds post-stimulus onset, with different feature types exerting independent effects. Thus, the effects of semantic features are clearly complex, modulated by feature type, and occur very early in the timecourse of lexical processing.

Unlike Rabovsky et al. (2012), other studies have found effects of number of associates on the ERP response. Müller et al. (2010) observed larger N400 components to items with a high number of associates (e.g., SPOON—FORK) than those with fewer associates in a lexical decision task. Laszlo and Federmeier (2011) likewise observed larger N400 amplitudes for items with many associates relative to those with fewer associates. The frequency of the associates also exerted an effect on the ERP waveform: items with higher frequency associates elicited larger N400 amplitudes than those with lower frequency associates.

Both of these studies found an *increase* in N400 amplitude for more or higher-frequency lexical associates, contrary to the prediction that items with more associates (i.e., with richer representations) should be easier to process and thus elicit less neural activation and smaller N400 amplitudes. Laszlo and Federmeier interpreted this finding as indicating that lexical associates, particularly those of high frequency, serve as better "lures" away from the target item. That is, high frequency words generally elicit a smaller amplitude N400 than low frequency words because they are easier to activate; in the case of high frequency associates, there is a greater tendency of the associate to become activated, luring activation from the target item and resulting in larger N400 amplitude. This interpretation is consistent with research suggesting that the N400 reflects inhibitory function (Debruille, 1998); the lexical associate competes with the target item and must be inhibited, leading to a larger N400. Such competition would not be expected in the case of items with multiple related senses, because the senses are not in competition for activation.

The present study extends previous research on ERP measures of semantic richness by examining ERP response to lexical items with high and low numbers of senses, an area that remains to be explored. Our previous research (Taler et al., 2009) found that lexical items with two related literal senses (e.g., CHICKEN meaning animal or meat) elicited smaller N400 amplitudes in healthy older adults than homonyms (i.e., items with two unrelated senses, such as BANK meaning "a financial institution" or "the side of a river"). However, that study did not assess the effect of number of senses, an important measure of semantic richness that has been demonstrated in behavioral studies to exert an effect on lexical processing. Specifically, Rodd et al. (2002) demonstrated facilitated processing for words with multiple related senses and a disadvantage for words with multiple meanings. Consistent with our previous research (Taler et al.), as well as previous behavioral research indicating faster processing of semantically rich items (Pexman et al., 2002, 2003, 2008) including multiple related senses (Rodd et al., 2002), we anticipate that high-sense words will elicit smaller N400 amplitudes and faster response times in a lexical decision task than low-sense words. Unlike lexical associates, related senses do not constitute a "lure" away from the target item, and hence we do not predict an increase in the amplitude of the N400 to high-sense items, as was observed for items with many lexical associates (Laszlo and Federmeier, 2011).

METHODS

PARTICIPANTS

Participants included 20 right handed native monolingual English speakers (11 females) recruited from the University of Ottawa community. Their mean age was 22 years ($SD = 1.8$) and at the time of testing they had completed an average of 16 years of education ($SD = 1.1$). Prior to the testing session all participants completed a self-report health and history questionnaire to ensure that they were in good health and did not suffer from any conditions and were not taking any medications that are known to affect cognitive function. All participants showed normal cognitive function as measured by the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) and had no neurological or psychiatric history. This study was approved by the Bruyère Research Institute and University of Ottawa research ethics committees.

MATERIALS AND APPARATUS

Stimuli

The experiment included a total of 134 stimuli in three conditions. The "high" condition comprised 32 stimuli with many related senses (e.g., EYE), while the "low" condition comprised 32 stimuli with few related senses (e.g., GYM)[1]. Number of senses was determined using WordNet (Princeton University, 2010). The experiment included 70 pseudowords matched to real word stimuli for length, orthographic neighborhood density (total number of orthographic neighbors, i.e., the N-metric), and bigram frequency by position, using data from the English Lexicon Project (Balota et al., 2002). The stimuli in the two real-word conditions (high and low) were matched for length and for bigram frequency by position using data from the English Lexicon Project, for frequency using log-transformed values from the CELEX database (Baayen et al., 1993), and for familiarity, concreteness, and imageability using data from the MRC Psycholinguistic Database (Coltheart, 1981). They were also balanced for number of lexical associates using the University of South Florida Free Association Norms (Nelson et al., 1998). Current databases do not provide sufficient numbers of items for control of number of features; however, as noted by Rabovsky et al. (2012), number of features and number of associates tend to be highly correlated. High-sense items were of higher orthographic neighborhood density than low-sense items ($p < 0.05$). Stimulus characteristics are provided in **Table 1**. All stimulus

[1]Six additional real-word items were not included in analyses; participants thus saw 140 stimuli in total.

Table 1 | Characteristics of experimental stimuli: mean (standard deviation).

	High-sense words	Low-sense words	Pseudowords
Number of related senses	6.72 (2.57)	1.70 (0.47)	N/A
Frequency (CELEX log-transformed)	1.32 (0.46)	1.15 (0.55)	N/A
Length	5.22 (1.41)	5.69 (1.35)	5.51 (1.34)
Orthographic neighborhood density	4.81 (4.58)	2.75 (3.34)	3.69 (3.63)
Bigram frequency by position	1340.59 (735.40)	1575.69 (837.66)	1396.66 (688.59)
Concreteness	561.59 (69.30)	571.97 (69.06)	N/A
Familiarity	539.34 (49.55)	530.53 (47.14)	N/A
Imageability	568.13 (52.05)	589.57 (47.69)	N/A
Number of associates	15.09 (5.58)	12.74 (5.40)	N/A

matching was done using independent sample t-tests (comparing high to low-sense items, and all real words to pseudowords).

Experimental task

Participants completed a lexical decision task, in which they decided if each stimulus was a real word in English or not, and indicated their response using the "a" and "l" keys on the keyboard; electrophysiological recording took place simultaneously. Stimuli were presented in white 18 point Courier New font on a black background for 2000 ms or until a response was detected. Prior to each stimulus, a fixation cross was presented for 500 ms at the center of the monitor and participants were asked to maintain their gaze on the fixation between the presentation of each word. The experiment was run using E-Prime 2.0 presentation software (Psychology Software Tools, Pittsburg, PA, USA). Stimuli were presented to participants on a Dell OptiPlex 780 desktop computer with Windows XP Professional operating system, an Intel Core 2 Duo processor and a 20″ monitor.

EEG recording

The continuous EEG was recorded from 32 electrode sites according to the international 10–20 system of electrode placement using tin electrodes and a commercially available nylon cap (Electro-Cap International, INC., Eaton, OH, USA). A cephalic site was used as the ground and all active sites were referenced online to linked ears. Four additional electrodes were used to record the horizontal and vertical electro-oculogram (EOG); the horizontal EOG was recorded from electrodes placed at the outer canthus of each eye and the vertical EOG from electrodes placed above and below the left eye. The EEG was amplified using NeuroScan NuAmps (NeuroScan, El Paso, TX, USA) and was sampled at a rate of 500 Hz in a DC to 100 Hz bandwidth. Electrical impedances were kept below 5 kΩ during EEG recording. The EEG data were processed offline using NeuroScan 4.3 EDIT software (NeuroScan, El Paso, TX, USA). We applied a 30 Hz lowpass filter, vertical EOG artefact was corrected using a spatial filter (NeuroScan EDIT 4.3), and trials containing horizontal EOG artefact exceeding $\pm 50\,\mu V$ were excluded from averaging, as were trials containing deflections exceeding $\pm 100\,\mu V$. The electrophysiological time epoch was 1100 ms comprised of a 100 ms pre-stimulus baseline and 1000 ms following the onset of the stimulus word. Averages were computed based on the three conditions of the experimental task and were baseline corrected to a $0\,\mu V$ average of the 100 ms pre-stimulus interval. Only correct trials were included in averages and all averages contained a minimum of 30 trials.

PROCEDURE

Participants were seated in a comfortable chair and informed consent was obtained. Given that this investigation formed part of another study, several neuropsychological tasks were performed prior to setting up the participant for EEG recording. These tasks took approximately 20 min to complete and were followed by the application of the electrodes for EEG recording, which took approximately 30 min. The experimental task examined here took approximately 5 min to complete and was followed by an additional task that was approximately 30 min long. In total the testing session ranged from 1.5 to 2 h with 30–45 min of EEG recording. Following the testing session the purpose of the experiment was explained in detail and any questions that the participant had were answered. Participants were compensated $10 per h of participation.

RESULTS

All statistical analyses were conducted using PASW Statistics v. 18. Only correct trials were included in the analyses.

BEHAVIOURAL RESULTS

Trials with RTs exceeding ±2.5 standard deviations from the mean were excluded as outliers, which resulted in the removal of a total of 3.9% of trials from the high sense condition (1.7% errors, 2.2% outliers) and 3.3% of trials from the low sense condition (1.4% errors, 1.9% outliers). The accuracy and reaction time (RT) data were analyzed in separate paired samples t-tests. There was an effect of number of senses on RT [$t_{(19)} = -2.07$, $p = 0.05$], but not on accuracy. Specifically, words with many senses elicited faster RTs ($M = 657.8$ ms, $SD = 137.5$) than those with few senses ($M = 680.0$ ms, $SD = 134.8$). In terms of accuracy, both conditions elicited few errors, with an accuracy rate of 98.3% ($SD = 2.4$) for the high sense condition and 98.6% ($SD = 1.9$) for the low sense condition.

ELECTROPHYSIOLOGICAL RESULTS

Two participants were excluded from analysis of the electrophysiological data due to technical difficulties during EEG recording; thus 18 participants were included in all analyses of the electrophysiological data. Given that we were interested in the N400, we analyzed sites where the N400 was largest (i.e., central and posterior sites) and only included the time interval that best represented the N400. Visual inspection indicated an early divergence in the waveforms; thus, we examined the 200–550 ms time window by subdividing it into consecutive 50 ms time bins (i.e., 200–250 ms, 250–300 ms, 300–350 ms, . . ., 500–550 ms). Time was then entered into the ANOVA as a within-subjects variable. We analyzed the midline, and the left and right lateral sites in three separate 2 (Condition: high vs. low) × 3 (Site: C4, CP4, P4 or Cz, CPz, Pz or C3, CP3, P3) × 7 (Time: 0–550 ms) repeated measures analyses of variance where the dependent variable was the average ERP amplitude in each of the 50 ms time bins. First we report the results from the midline sites, followed by the results from the lateral sites.

Grand averages waveforms are depicted in **Figure 1**. Analysis of the midline sites revealed a main effect of Condition, $F_{(1,\ 17)} = 5.09$, $MSE = 57.35$, $p = 0.04$, indicating a larger amplitude N400 for words with few related senses than words with many related senses. This finding was further supported by the analysis of the lateral sites. Right lateral sites demonstrated a similar effect of Condition [$F_{(1,\ 17)} = 6.92$, $MSE = 48.20$, $p = 0.02$], with words with few related senses demonstrating a larger N400 amplitude than words with many related senses. Finally, analysis of the left lateral sites revealed a trend toward a main effect of Condition, $F_{(1,\ 17)} = 3.53$, $MSE = 64.97$, $p = 0.08$, again revealing larger N400 amplitudes for words with few related senses relative to words with many related senses. There were no meaningful significant interactions in any of the analyses performed. We note that high-sense items were also of higher orthographic neighborhood density than low-sense items; however, the effect (larger N400 amplitudes for low- than high-sense items) is in the opposite direction to the reported effects of neighborhood density (larger N400 amplitudes for high- than low-neighborhood density items) (Holcomb et al., 2002).

DISCUSSION

The present study examined behavioral and ERP response to words with high and low levels of semantic richness, as measured by the number of related senses that the word possesses. Participants responded more quickly to high- than low-sense items, consistent with previous research (Rodd et al., 2002; Yap et al., 2012). We found smaller N400 amplitudes to high- than low-sense words in midline and right lateral sites, with the waveforms for the two conditions diverging at around 200 milliseconds post-stimulus onset.

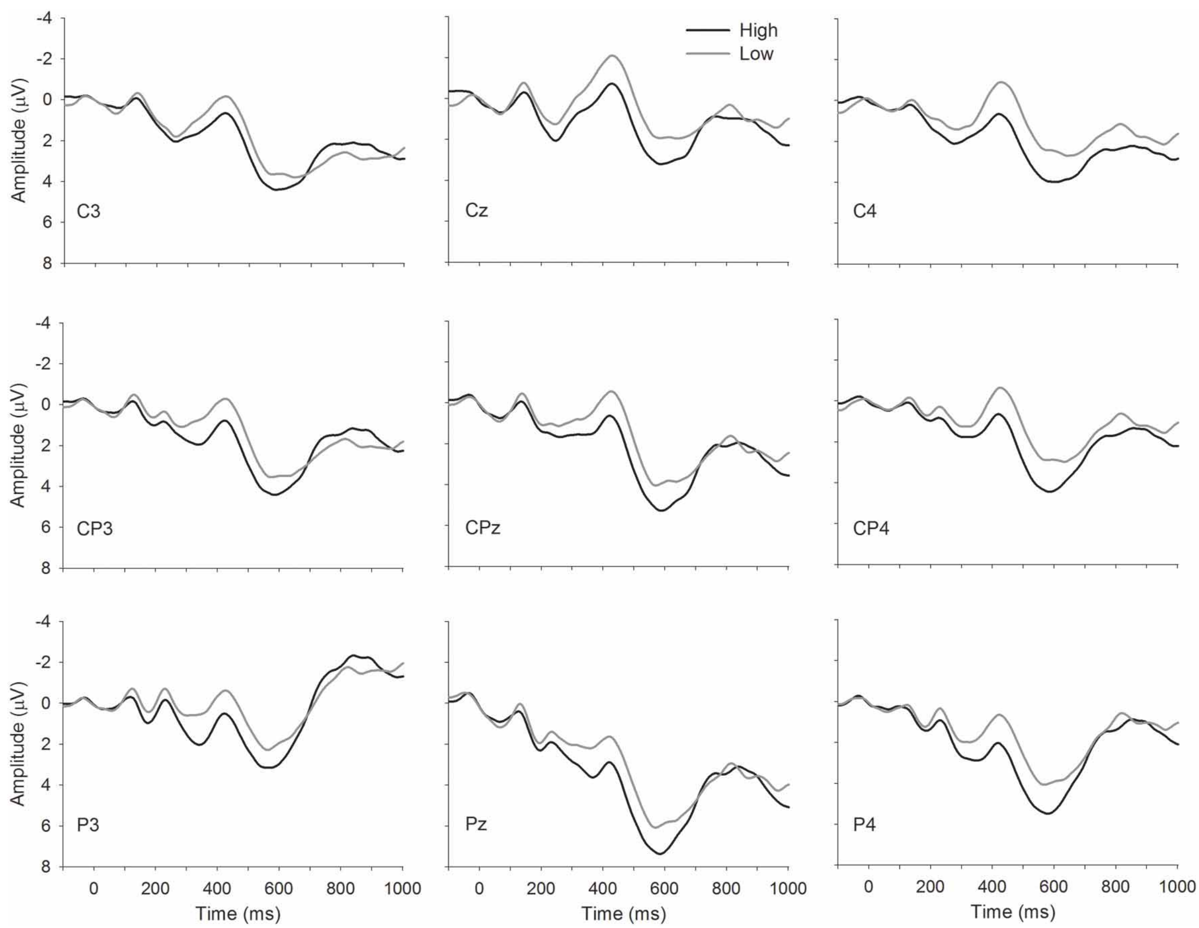


FIGURE 1 | Grand average waveforms showing high- and low-sense items. Time is plotted on the *x*-axis and amplitude in microvolts on the *y*-axis; following convention, negative amplitudes are plotted upwards.

These results indicate that readers activate information about the number of senses a word possesses very early in processing, consistent with previous research on number of semantic features (Amsel, 2011; Rabovsky et al., 2012). However, interestingly, the direction of the effect differs between our findings and Rabovsky et al., who found a larger N400 to high-feature than low-feature items. Similarly, Amsel (2011) found more negativity in the N400 window to high-feature items. Laszlo and Federmeier (2011) also found larger N400s to words with more or higher-frequency lexical associates than fewer or lower-frequency associates. Rabovsky et al. do not offer a theoretical account for their findings, but Laszlo and Federmeier suggest that lexical associates, particularly those of high frequency, may serve as better "lures" for the target item, thus invoking an account similar to that of Debruille (1998), who suggests that the N400 may index inhibitory function. Given that our stimuli differ in the number of related senses and not number of associates, it is not surprising that they do not invoke inhibition. That is, the multiple senses of our stimuli are related and therefore activation of a larger number of senses should facilitate processing because there is no need to inhibit any of the senses, as is the case when an associate is activated. Thus, stimuli with a high number of related senses should elicit faster RTs and smaller amplitude N400s, as the results demonstrate.

These results also contrast with the findings reported in the literature that concrete words elicit larger N400 amplitudes than abstract words (Kounios and Holcomb, 1994; Holcomb et al., 1999; West and Holcomb, 2000), possibly reflecting higher semantic richness associated with high concreteness [although note that Kounios et al. (2009) suggest that previous studies of concreteness may not have adequately controlled for relevant semantic variables]. Similarly, Rabovsky et al. (2012) recently reported effects of newly acquired semantic information: newly-learned words associated with more semantic information showed larger amplitude N400s than those associated with less information.

The findings are, however, consistent with our previous research (Taler et al., 2009), which found smaller N400 components to metonyms (lexically ambiguous items with two related literal senses, such as CHICKEN) than to homonyms (lexically ambiguous items with two unrelated meanings). We argue that metonyms are semantically richer than homonyms, which are associated with two unrelated—and thus competing—lexical entries rather than a single lexical entry or multiple closely-linked entries. This finding indicates that relatedness of senses impacts the ERP response, with reduced N400 amplitude associated with related relative to unrelated senses. Similarly, Rodd et al. (2002) differentiate between related and unrelated senses such that processing is facilitated for words with multiple related senses, but not for words with multiple unrelated senses. Our findings are also in line with behavioral evidence indicating shorter response latencies to high- than low-sense items (Rodd et al., 2002; Yap et al., 2012), as well as with fMRI evidence showing less neural activation to high- than low-number of associate lexical items (Pexman et al., 2007). It should be noted, however, that N400 amplitude does not map perfectly onto response latency; shorter response times to items that elicit larger N400 responses have been reported in the case of orthographic neighborhood density (Holcomb et al., 2002), number of lexical associates (Müller et al., 2010), and concreteness (West and Holcomb, 2000).

In sum, the varied findings reported in the literature indicate that semantic richness effects are far from straightforward. One possibility that arises from this discrepancy is that different measures of semantic richness may exert different influences on lexical processing, a conclusion also reached by Rabovsky et al. (2012). This hypothesis has been explored by Pexman et al. (2008), who found that three measures of semantic richness (number of features, number of associates, and contextual dispersion) accounted for unique variance in a lexical decision task. Similarly, Yap et al. (2012) found that different types of semantic richness exerted different effects in various word recognition tasks, suggesting that various types of semantic information are used adaptively depending upon task demands. We suggest that some types of semantic richness (such as number of lexical associates) may require greater inhibitory control than others (such as number of related senses); this hypothesis is supported by opposite effects on the N400 component of these two semantic factors.

In the present results, we interpret the reduced N400 amplitude to high- relative to low-sense words to reflect reduced processing demands in the former case. This interpretation is supported not only by existing literature (Yap et al., 2012), but also by the shorter response latencies to high-sense words than to low-sense words. The present study focused exclusively on the effect of number of related senses on lexical decision latency and ERP response. Future research should further explore different measures of semantic richness in combination with number of senses, as these measures likely exert differential effects on the ERP response, and may indeed interact. We also note that, although high- and low-sense items did not differ significantly in terms of frequency, high-sense items were numerically higher in frequency than low-sense items, which may have influenced the present results.

In summary, we demonstrated reduced N400 amplitude and shorter response latencies to high-than low-sense lexical items. These results indicate greater ease of processing of words with many related senses relative to words with fewer senses. The ERP waveform began to diverge between the two conditions as early as 200 milliseconds post-stimulus onset, indicating that this semantic information is accessed very early in lexical processing. These results shed further light on the organization of semantic information in the lexicon, as well as the timecourse of lexical access.

ACKNOWLEDGMENTS

This research was supported by a Catalyst Grant awarded to Vanessa Taler and Shanna Kousaie by the Canadian Institutes of Health Research. We would like to thank Chloé Corbeil, Christianne Laliberté, Jessica Kumar, Jenna Schultz, and Christine Sheppard for assistance in data collection and processing.

REFERENCES

Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia* 49, 970–983.

Baayen, R. H., Piepenbrock, R., and van Rijn, H. (1993). *The CELEX Lexical Database (Release 1) [CD-ROM]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania (Distributor).

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., et al. (2002). *The English Lexicon Project: A Web-Based Repository of Descriptive and Behavioral Measures for 40,481 English Words and Nonwords.* (Electronic database). Available online at: http://elexicon.wustl.edu/

Coltheart, M. (1981). The MRC psycholinguistic database. *Q. J. Exp. Psychol.* 33, 497–505.

Debruille, J. B. (1998). Knowledge inhibition and N400: a study with words that look like common words. *Brain Lang.* 62, 202–220.

Holcomb, P. J., Grainger, J., and O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *J. Cogn. Neurosci.* 14, 938–950.

Holcomb, P. J., Kounios, J., Anderson, J. E., and West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 721–742.

Kounios, J. (1996). On the continuity of thought and the representation of knowledge: electrophysiological and behavioral time-course measures reveal the levels of structure in semantic memory. *Psychon. Bull. Rev.* 3, 265–286.

Kounios, J., Green, D. L., Payne, L., Fleck, J. I., Grondin, R., and McRae, K. (2009). Semantic richness and the activation of concepts in semantic memory: evidence from event-related potentials. *Brain Res.* 1282, 95–102.

Kounios, J., and Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 804–823.

Kutas, M., and van Petten, C. K. (1994). "Psycholinguistics electrified: event-related brain potential investigations," in *Handbook of Psycholinguistics,* ed M. A. Gernsbacher (San Diego, CA: Academic Press, Inc.), 83–143.

Laszlo, S., and Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology* 48, 176–186.

Müller, O., Duñabeitia, J. A., and Carreiras, M. (2010). Orthographic and associative neighborhood density effects: what is shared, what is different? *Psychophysiology* 47, 455–466.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal cognitive assessment: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). *The University of South Florida Word Association, Rhyme, and Word Fragment Norms*. Available online at: http://www.usf.edu/FreeAssociation/

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., and Goodyear, B. C. (2007). The neural consequences of semantic richness. *Psychol. Sci.* 18, 401–406.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on visual word recognition. *Psychon. Bull. Rev.* 15, 161–167.

Pexman, P. M., Holyk, G. G., and Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Mem. Cognit.* 31, 842–855.

Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of semantics in visual word recognition: number of features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.

Princeton University. (2010). *About WordNet, 2012*. Available online at: http://wordnet.princeton.edu

Rabovsky, M., Sommer, W., and Abdel Rahman, R. (2012). The time course of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:11. doi: 10.3389/fnhum.2012.00011

Rodd, J. M., Gaskell, G., and Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: semantic competition in lexical access. *J. Mem. Lang.* 46, 245–266.

Taler, V., Klepousniotou, E., and Phillips, N. A. (2009). Comprehension of lexical ambiguity in mild cognitive impairment and Alzheimer's disease. *Neuropsychologia* 47, 1332–1343.

West, W. C., and Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *J. Cogn. Neurosci.* 12, 1024–1037.

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., and Huff, M. J. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.*6:72. doi: 10.3389/fnhum.2012.00072

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 September 2012; accepted: 07 January 2013; published online: 31 January 2013.

Citation: Taler V, Kousaie S and López Zunini R (2013) ERP measures of semantic richness: the case of multiple senses. Front. Hum. Neurosci. 7:5. doi: 10.3389/fnhum.2013.00005

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 14 August 2012
doi: 10.3389/fnhum.2012.00234

Does richness lose its luster? Effects of extensive practice on semantic richness in visual word recognition

Ian S. Hargreaves* and Penny M. Pexman*

Department of Psychology, University of Calgary, Calgary, AB, Canada

Edited by:
Melvin Yap, National University of Singapore, Singapore

Reviewed by:
Emmanuel Keuleers, Ghent University, Belgium
Daniel Mirman, Moss Rehabilitation Research Institute, USA

****Correspondence:***
Ian S. Hargreaves and Penny M. Pexman, Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada.
e-mail: ishargre@ucalgary.ca; pexman@ucalgary.ca

Previous studies have reported facilitatory effects of *semantic richness* on word recognition (e.g., Yap et al., 2012). These effects suggest that word meaning is an important contributor to lexical decision task (LDT) performance, but what are the effects of repeated LDT practice on these semantic contributions? The current study utilized data from the British Lexicon Project (BLP) in which 78 participants made lexical decision judgments for 28,730 words over 16 h. We used linear mixed effects to detect practice-driven changes in the explanatory variance accounted for by a set of lexical predictors that included numerous indices of relative semantic richness, including imageability, the number of senses and average radius of co-occurrence (ARC). Results showed that practice was associated with decreasing effects of predictors such as word frequency and imageability. In contrast, ARC effects were only slightly diminished with repeated practice, and effects of the number of senses and the age of acquisition were unaffected by practice. We interpret our results within a framework in which variables may dynamically influence lexical processing and the post-lexical decision making mechanisms that also contribute to LDT performance.

Keywords: lexical decision task, practice effects, semantic richness, visual word recognition, reaction time

INTRODUCTION

Over the past several decades, considerable research attention has been devoted to the study of visual word recognition. As a result, there are now a number of well-established findings in the word recognition literature, and consistent behavioral findings have generally been interpreted as evidence for the stable underlying representational structure of the word recognition system. For instance, consistent behavioral effects of different word characteristics, such as word length effects (faster lexical decisions for shorter words, e.g., New et al., 2006), frequency effects (faster lexical decisions for words that appear more frequently in language, e.g., Balota et al., 2004), and semantic richness effects (faster lexical decisions to words associated with more semantic information, e.g., Pexman et al., 2008) have fuelled assumptions about contributions made by these kinds of information to the process of recognizing words. Indeed, these are among the standard word recognition effects that all models of word recognition are designed to explain.

What is potentially more challenging for models to accommodate, however, is the possibility that there is variability in the process as a function of context or experience, and it is this variability that is the focus of the present work. There is evidence that visual word recognition is a dynamic process; participants can make adjustments to this process in order to optimize performance under various task conditions. For instance, in the standard version of the lexical decision task (LDT), the non-word stimuli are orthographically similar to words (word-like letter combinations, e.g., SLINT) but do not share the sound of a real word if pronounced. If the non-words are made more similar to words, for example by using pseudohomophones (non-words that do sound like real words if pronounced, e.g., BRANE), then several changes in LDT performance can be observed: latencies are slower for both word and non-word responses, and certain behavioral effects (e.g., the word frequency effect) are reliably larger (e.g., Stone and Van Orden, 1993; Lupker and Pexman, 2010).

Similarly, the structural overlap between non-words and words has been shown to create a systematic bias in LDT responses. Keuleers and Brysbaert (2011) designed an algorithm capable of successfully predicting the likelihood of generating a "word" or "non-word" response based solely on the structural similarity of the current trial to past trials (whether "word" or "non-word"). They found that the choice of non-words could bias responses; when non-words were generated from real words (e.g., by manually changing one or two letters of real words as in the English Lexicon Project-ELP; Balota et al., 2007) the high degree of similarity between words and non-words led to a counterintuitive bias to respond "word" when presented with a non-word (and vice versa). This bias also predicted behavioral slowing for both "word" and "non-word" responses in the ELP data and could be mitigated by reducing the structural similarity between word and non-words (e.g., by using the Wuggy algorithm to generate non-words; Keuleers and Brysbaert, 2010). This suggests that participants are implicitly tracking systematic trends in the structural properties of items in order to optimize decision-making in the LDT.

Further, Kiefer and Martens (2010) showed that even purportedly unconscious effects in word recognition can be modulated by context. That is, Kiefer and Martens examined masked semantic priming effects in LDT, involving faster latencies and attenuation of the N400 ERP component for related (table-chair) compared to unrelated (car-hen) targets. Context was manipulated by a perceptual induction task that required directing attention to either semantic or perceptual features. Results showed that semantic priming (in both behavior and ERPs) was enhanced following a semantic induction task and was attenuated following a perceptual induction task. These authors interpreted their results as consistent with the attentional sensitization model, by which top–down mechanisms enhance or attenuate different processing streams in order to facilitate processing that is compatible with higher level goals.

There is also evidence that the word recognition process is shaped by a reader's lexical experience. The effects of practice or experience on word recognition have typically been studied between-subjects, by comparing word recognition behavior in individuals who differ on some experience dimension. For instance, Yap et al. (2011) examined the relationship between readers' vocabulary knowledge and their word recognition behavior, using trial-level LDT data from the ELP megastudy (3374 LDT trials for each of 819 participants). Yap et al. reported that readers with higher vocabulary scores were faster to respond, and that higher vocabulary scores were associated with smaller frequency and semantic effects. Once individual differences in processing speed were controlled, however, Yap et al. found that the relationship between vocabulary scores and a composite frequency/semantic measure (comprised of frequency, number of senses, and also semantic neighborhood density) was only marginally significant.

A small number of studies have explored the effects of experience on lexical processing by examining the word recognition performance of Scrabble experts (Halpern and Wai, 2007; Tuffiash et al., 2007; Hargreaves et al., 2012). Scrabble experience provides the opportunity to develop strong word recognition skills. Since, the ability to detect phony plays (when an opponent plays a non-word) in Scrabble is essential to competitive success, Scrabble players develop extensive knowledge of the lexical status of different letter strings. Hargreaves et al. (2012) showed that this knowledge was associated with faster responses and smaller concreteness effects in LDT, and interpreted this to mean that Scrabble experts showed less reliance on the meanings of words in order to judge lexicality. The behavior of these visual word recognition experts highlights the experience-driven nature of visual word recognition. However, although expert and novice groups were matched on a number of related variables (e.g., vocabulary size, exposure to print, and education), the between-subjects nature of this work means that there is always the possibility that the group differences in word recognition behavior are not due to Scrabble experience but rather to some other uncontrolled group difference.

A within-subjects approach to the study of practice effects in lexical processing was adopted in a recent megastudy. Keuleers et al. (2010) examined the effects of practice on word recognition within subjects by comparing performance across 57 blocks of a LDT using 14,089 Dutch words. The authors found that over time, effects of word frequency diminished with repeated practice in the LDT. Interestingly, the influence of practice on effects of word length and the mean Levenshtein distance to the nearest 20 orthographic neighbors was less clear, as neither formed any linear relationship with repeated LDT practice. That repeated practice selectively influenced some, but not other, lexical variables suggests that practice in the LDT may influence the decision on many levels. Indeed, in another megastudy Dutilh et al. (2009) identified numerous influences of practice on decision-making in the LDT. Dutilh et al. examined the effects of practice on word recognition within subjects by comparing performance across the 25 blocks in a 10,000 trial lexical decision study. They reported that practice was associated with faster and less variable response latencies. Further, diffusion model analyses (e.g., Ratcliff et al., 2006; Ratcliff and McKoon, 2008) showed an increased rate of information processing (increased drift rate), decrease in response caution (or narrowing of decision boundaries), and decrease in time required for common processes executed irrespective of the decision (decreased non-decision time) with practice. Non-decision time facilitation was attributed to increased familiarity with the task demands. These results suggest that with extensive practice participants modify lexical decision making processes in order to optimize performance.

THE PRESENT STUDY

In the present work we also adopted a within-subjects approach to the study of practice effects in LDT, capitalizing on a recent LDT megastudy of 28,730 trials known as the British Lexicon Project (BLP; Keuleers et al., 2011). Each participant in the BLP made lexical decisions to 14,365 words (and the same number of non-words) over 16 h of testing. As such, the BLP provides the largest dataset currently available with which to examine effects of practice. We anticipated that, as in the Keuleers et al. (2010) and Dutilh et al. (2009) studies, participants would become much faster across response blocks. Indeed, as Keuleers et al. (2011) noted, the practice effect for word trials in the BLP study was around 100 ms. Our particular interest was in whether participants would show changes in their weighting or reliance on different types of lexical and semantic information as they became more and more practiced at making lexical decisions.

To assess this question, we examined participant behavior across blocks of trials (each block included 500 trials) in the BLP. We assessed the extent to which behavior across blocks could be predicted by orthographic variables, including length and orthographic neighborhood characteristics, word frequency (Brysbaert and New, 2009), and semantic richness. The semantic richness dimension we examined in our analysis of the entire dataset was ARC (average radius of co-occurrence). This measure of semantic neighborhood density was developed by Shaoul and Westbury (2006, 2010) and was derived from the HAL model of lexical co-occurrence (Burgess, 1988; Burgess and Lund, 2000). ARC is based on the average distance of a target word to its neighbors (within a threshold) in high-dimensional semantic space. Shaoul and Westbury (2010) reported that LDT latencies were related to ARC values, such that latencies were faster for words with denser semantic neighborhoods.

MATERIALS AND METHODS

DEPENDENT MEASURE

Lexical decision data were obtained from the BLP (http://crr.ugent.be/blp/), an online database containing trial-level LDT data for 28,730 monosyllabic and disyllabic words (Keuleers et al., 2011). A full description of the methodology used in collecting the LDT data is available in Keuleers et al. (2011). For the present analysis it is worth noting that BLP participants were instructed to attempt to maintain a consistently high (80%) level of accuracy, a criterion that was made challenging by the inclusion of a large number of low-frequency words (as low as 0.02 per million words). Participants were instructed to try to keep their average RT below 1 s, however, trials did not time out if they exceeded this value. In the subsequent analysis we included only correct responses for words, further, we required these responses to fall between 200 and 1700 ms. Respectively, these criteria excluded 24.3% and <2% of the data.

ANALYSIS—FULL BLP DATASET

From the original set of 28,730 words, we were able to obtain a complete set of predictors for 25,463 words, and subsequent analysis proceeded with these items. We constructed a linear-mixed effects (LME) model, implemented using the lme4 library (version 0.999375-42; Bates, 2007) in R (version 2.14.2, Bates, 2007; Baayen et al., 2008; R Development Core Team, 2010). Participants and items were treated as random factors. We used an iterative model fitting procedure included in the package *LMERConvenienceFunctions* (Tremblay, 2011) to generate a random effects structure that provided the best combination of goodness of fit and parsimony (i.e., number of parameters), as determined through iterative testing using likelihood ratio tests. In order to characterize the contributions of semantic richness to LDT performance, ARC values (continuous)[1] were included as a fixed effects factor. We controlled for additional variance by including letter length (continuous), the orthographic Levenshtein distance to the nearest 20 neighbors (OLD20; continuous; Yarkoni et al., 2008), and log transformed frequency in the SUBTLEX corpus (continuous; Brysbaert and New, 2009), as fixed factors. In addition, we controlled for reaction time on the previous trial by including it as a fixed factor (continuous). All predictors were centered to reduce collinearity (**Table 1**). Finally, to assess the presence of practice effects we included an interaction term between block (continuous) and all other fixed factors. The significance of individual fixed effects parameters was assessed by subtracting the number of fixed effects in the model (12) from the number of observations (773,527). With hundreds of thousands of degrees of freedom the *t* distribution approximates the normal distribution, thus this approach provides a reasonably conservative estimate of statistical significance (Baayen, 2008). Although, block was included as a continuous fixed factor, in order to facilitate the interpretation of any significant interactions between block and another fixed factor we subsequently dichotomized block using a median split. Thus, block was divided into early (less than block 28) and late (greater than block 28 but excluding the shortened block 57) epochs, and a reduced model without the interaction of block was fitted to the data from early and late blocks.

RESULTS—FULL BLP DATASET

The results of the LME modeling are presented in **Table 2**. We observed longer reaction times when the previous trial's reaction time was longer. In addition, responses were slower when words had greater mean Levenshtein distance to their nearest 20 neighbors. We observed significant facilitation for higher frequency words and words with denser semantic neighborhoods. In addition, there was a main effect of practice, with mean reaction times falling as participants gained more exposure to the LDT.

Importantly, all but one (OLD20) of our main effects were qualified by a significant interaction with block, indicating that the sizes of the lexical and semantic effects were modulated by practice. As displayed in **Figure 1**, follow-up analyses using a median split of block revealed that the size of length, frequency and semantic neighborhood density (ARC) effects were all reduced in later blocks relative to earlier blocks. These results suggest that participants were reweighting their reliance on different types of lexical and semantic information as they became more and more practiced at making lexical decisions.

While the observation of practice effects is interesting, past investigations of LDT megastudies (e.g., Keuleers et al., 2010) have revealed similar effects of practice on the LDT. Although, the absolute size of this reduction is quite modest, the observation that semantic richness effects are also modulated by practice is a novel finding that raises a question about whether other descriptions of relative semantic richness would show similar practice effects[2]. Numerous studies have revealed that the changes

[1] http://www.psych.ualberta.ca/~westburylab/downloads/westburylab.arcs.ncounts.html

[2] Our thanks to editor for this suggestion.

Table 1 | Correlations between centered predictor variables and dependent variable for 25,463 words.

Variable	1	2	3	4	5	6
1. Previous trial RT	–					
2. No. of letters	−0.00**	–				
3. Orthographic Levenshtein distance (Yarkoni et al., 2008)	−0.00**	0.76***	–			
4. Log frequency (Brysbaert and New, 2009)	0.00***	−0.37***	−0.29***	–		
5. ARC (Shaoul and Westbury, 2010)	0.00**	−0.28***	−0.19***	0.68***	–	
6. LDT RT	0.25***	0.11***	0.11***	−0.25***	−0.21***	–

***$p < 0.001$, **$p < 0.01$.

Table 2 | Effect sizes (bs), standard errors (SEs), and t values for linear mixed effects models of lexical decision reaction times to 25,463 words.

Fixed effects	Overall			Early			Late		
	b	SE	t	b	SE	t	b	SE	t
Previous trial RT	0.07	9.23	73.07*	0.08	0.00	95.13*	0.08	0.00	87.81*
Length	−8.58	0.77	−11.14*	−6.75	0.75	−8.98*	−5.99	0.68	−8.82*
OLD20	25.75	2.12	12.18*	22.92	2.09	10.95*	22.94	1.92	11.96*
Frequency	−54.64	1.35	−40.39*	−47.84	1.45	−33.08*	−42.79	1.37	−31.32*
ARC	−119.10	6.98	−17.05*	−113.20	6.64	−17.05*	−106.80	6.39	−16.70*
Block	−1.17	0.16	−7.44*						
Block × previous RT	0.0003	0.00	8.47*						
Block × length	0.05	0.01	4.25*						
Block × OLD20	−0.02	0.03	−0.61						
Block × frequency	0.36	0.02	21.37*						
Block × ARC	0.30	0.10	3.04*						

Random effects		Variance (Overall)	Variance (Early)	Variance (Late)
Subject	(Intercept)	6615.7	5315.94	5678.35
	Block	1.93	N/A	N/A
	Length	25.28	30.46	24.38
	OLD20	216.37	255.56	213.29
	Frequency	96.71	133.31	120.20
	ARC	2200.60	2397.46	2304.71
Item	(Intercept)	2211.5	2566.56	2128.60
	Block	0.20	N/A	N/A
Residual		25171.10	27092.25	24607.13

*Note: *Indicates that the t-value was significant at $p < 0.05$.*

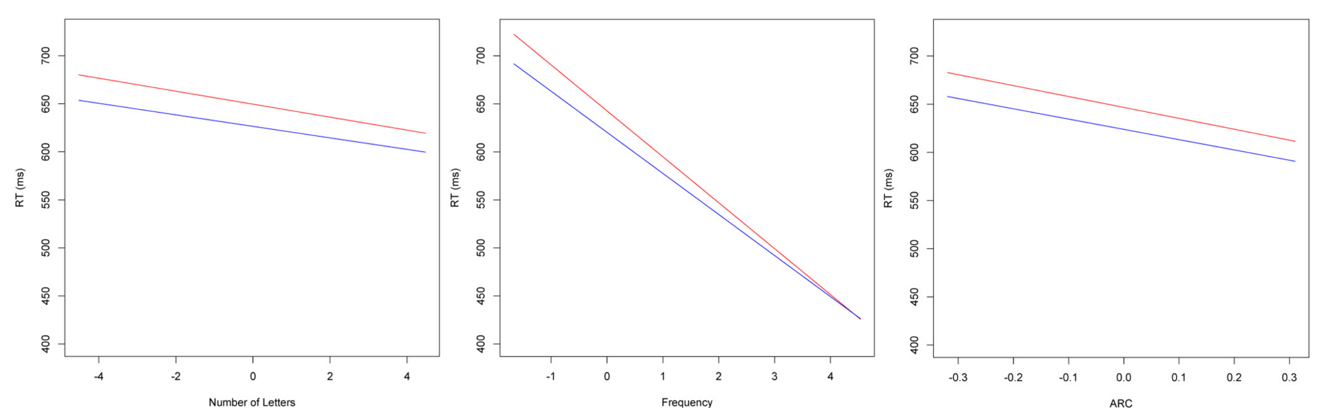


FIGURE 1 | Plots of fixed effects from a model fitted to LDT reaction times to 25,463 words for early (less than block 28; red) and late (greater than block 28 but excluding the shortened block 57; blue) blocks. Note: Length, number of letters (centered); Frequency, log SUBTLEX frequency (centered); ARC, semantic neighborhood density (centered).

in processing elicited by different tasks can lead to the selective recruitment of different descriptions of semantic richness, with some dimensions contributing to some tasks and not to others (e.g., imageability, the number of senses, ARC; Pexman et al., 2008; Yap et al., 2011, 2012). Similarly, the increased efficiency in the LDT that is purchased with practice may selectively influence some forms of semantic richness but not others. In order to assess this question we performed a separate analysis on a subset of items for which we had a complete set of predictors including several forms of semantic richness: the number of senses a word has (as evidenced by the number of discrete Wordsmyth entries; http://www.wordsmyth.net), a word's rated imageability, and finally a word's semantic neighborhood density as measured by ARC. In addition, we were able to introduce an additional control for the words' estimated age of acquisition.

ANALYSIS—RESTRICTED BLP DATASET

From the original set of 28,730 words, we were able to obtain a complete set of predictors for 3723 words, and subsequent analysis proceeded with these items. We constructed a LME model, again implemented using the lme4 library. Participants and items were treated as random factors. We used an iterative model fitting procedure included in the package *LMERConvenienceFunctions* (Tremblay, 2011) to generate a random effects structure that provided the best combination of goodness of fit and parsimony (i.e., number of parameters), as determined through iterative testing using likelihood ratio tests. The number of senses (continuous; Wordsmyth), imageability (continuous; Cortese and Fugett, 2004; Stadthagen-Gonzalez and Davis, 2006), and ARC values (continuous)[3] were included as fixed semantic richness factors. In order to control for additional variance we included letter length (continuous), the orthographic Levenstein distance to the nearest 20 neighbors (OLD20; continuous; Yarkoni et al., 2008), log transformed frequency in the SUBTLEX corpus (continuous; Brysbaert and New, 2009) and age of acquisition (AoA; continuous; Stadthagen-Gonzalez and Davis, 2006; Cortese and Khanna, 2008) as control variables. Finally, we controlled for reaction time on the previous trial by including it as a fixed factor (continuous). All predictors were centered to reduce collinearity (**Table 3**). In order to assess the presence of practice effects we included an interaction term between block (continuous) and all other fixed factors. Again, the significance of individual fixed effects parameters was assessed by subtracting the number of fixed effects in the model (18) from the number of observations (129,925; Baayen, 2008). In order to facilitate the interpretation of any significant interactions between block and another fixed factor we subsequently dichotomized block using a median split. Thus, block was divided into early (less than block 28) and late (greater than block 28 but excluding the shortened block 57) epochs, and a reduced model without the interaction of block was fitted to the data from early and late blocks.

[3] http://www.psych.ualberta.ca/~westburylab/downloads/westburylab.arcs.ncounts.html

RESULTS—RESTRICTED BLP DATASET

The results of the LME modeling are presented in **Table 4** and in **Figure 2**. Length was the only predictor that failed to form a significant relationship with reaction time. We observed facilitatory effects of block and frequency as increasing practice and word frequency led to shorter reaction times. Curiously, the effect of OLD20 ran in the opposite direction as that observed with the larger set of items. Words with less dense orthographic neighborhoods showed faster reaction times. In their analysis of the BLP data, Keuleers and colleagues (2010) were unable to find reliable effects of the related construct number of orthographic neighbors. Indeed, our follow-up to the significant interaction suggests that the influence of orthographic neighbors for this set of items may be highly variable, reaching significance in early blocks but not in later blocks. More expected was the observation that words learned later in life (as assessed by AoA ratings) had longer reaction times than those learned earlier in life. All semantic richness variables led to significant facilitation in the expected direction, with more senses, higher imageability ratings and denser semantic neighborhoods all being associated with faster responses in the LDT.

The main effects of orthographic Levenshtein distance, frequency, imageability and ARC were all qualified by significant interactions with block, indicating that only these variables showed a practice effect. As displayed in **Figure 3**, follow-up analyses using a median split of block revealed that the sizes of all of these effects were attenuated as participants gained more practice in the LDT. Interestingly, the decreases in the sizes of the effects for orthographic Levenshtein distance, frequency, and imageability with practice were much larger than that observed for the decrease in ARC effects with practice.

DISCUSSION

The results of the current investigation provide further evidence that extensive practice with the LDT leads to significant facilitation as participants' responses became faster with increasing experience with the LDT. The current results also provide an early look at the influence of practice effects on numerous lexical and semantic dimensions, and indicate that practice-driven optimization of processing in the LDT has a diverse set of

Table 3 | Correlations between centered predictor variables and dependent variable for 3723 words.

Variable	1	2	3	4	5	6	7	8	9
1. Previous trial RT	–								
2. No. of letters	−0.01	–							
3. Orthographic Levenshtein distance (Yarkoni et al., 2008)	−0.01*	0.74***	–						
4. Log frequency (Brysbaert and New, 2009)	0.00	−0.25***	−0.20***	–					
5. Age of acquisition	−0.00	0.22***	0.21***	−0.69***	–				
6. Imageability	0.00	−0.11***	−0.12***	0.13***	−0.48***	–			
7. Number of senses	0.00	−0.15***	−0.19***	0.41***	−0.37***	0.13***	–		
8. ARC (Shaoul and Westbury, 2010)	−0.00	−0.09***	−0.07***	0.74***	−0.48***	0.15***	0.33***	–	
9. LDT RT	0.29***	0.02***	0.00	−0.10**	0.10***	−0.05***	−0.06***	−0.09**	–

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

Table 4 | Effect sizes (bs), standard errors (SEs), and t values for linear mixed effects models of lexical decision reaction times to 3723 words.

Fixed effects	Overall			Early			Late		
	b	SE	t	b	SE	t	b	SE	t
Previous trial RT	0.06	0.00	26.25*	0.08	0.00	40.15*	0.09	0.00	41.13*
Length	−0.71	1.56	−0.46	−1.34	1.45	−0.92	−2.14	1.19	−1.80
OLD20	−15.25	3.46	−4.41*	−9.92	2.92	−3.39*	−3.49	2.60	−1.34
Frequency	−17.11	2.29	−7.47*	−15.30	2.05	−7.45*	−9.36	1.85	−5.07*
Imageability	−7.41	0.90	−8.21*	−6.24	0.81	−7.71*	−2.99	0.69	−4.35*
AoA	18.77	1.38	13.63*	17.01	1.13	15.04*	17.33	1.02	17.08*
Number of senses	−1.26	0.24	−5.29*	−1.18	0.19	−6.24*	−1.28	0.17	−7.64*
ARC	−156.60	13.34	−11.74*	−142.10	12.51	−11.36*	−126.40	11.15	−11.34*
Block	−0.89	0.15	−6.08*						
Block × previous RT	0.00	0.00	7.91*						
Block × length	−0.04	0.04	−1.09						
Block × OLD20	0.30	0.09	3.32*						
Block × frequency	0.17	0.05	3.12*						
Block × imageability	0.10	0.02	4.51*						
Block × AoA	−0.05	0.04	−1.41						
Block × number of Senses	0.00	0.01	0.26						
Block × ARC	0.71	0.28	2.54*						

Random effects		Variance	Variance	Variance
Subject	(Intercept)	4476.37	4495.0	5132.76
	Block	1.62	N/A	N/A
	Length	42.49	59.57	30.59
	Frequency	72.99	90.98	82.52
	Imageability	11.18	16.47	9.18
	ARC	4277.40	5078.81	4135.98
Item	(Intercept)	816.31	870.01	642.62
	Length	20.97	59.14	2.88
	Frequency	331.25	549.31	279.25
Residual		23358.14	24852.09	22290.12

*Note: * Indicates that the t-value was significant at $p < 0.05$.*

consequences. While practice led to clear attenuation of the contributions of some lexical and semantic dimensions (e.g., word frequency and imageability), for other dimensions we observed practice-driven attenuation that, though reaching statistical significance, had limited practical significance (e.g., ARC). Further, for the contributions of other dimensions we observed no appreciable practice-driven attenuation at all (e.g., the number of senses and AoA).

Performance in the LDT is thought to provide a window into the mechanisms that drive lexical processing, but it is also thought to depend upon the contributions of post-lexical decision-making mechanisms (Balota and Chumbley, 1984; Yap et al., 2011). Recent research investigating the influence of repeated practice in the LDT suggests that practice influences both lexical and post-lexical processes (Dutilh et al., 2009). Appealing to the diffusion model, Dutilh and colleagues argued that the influence of practice is observed not only in changes to the speed of information processing (as indexed by drift rate), but also in changes in the overall familiarity with the task demands, which allows participants to dynamically adjust their criteria for what evidence counts as a "word" or "non-word" response. Though interesting, these findings are framed in terms of the specific parameters of the diffusion model, parameters that are not transparently linked to any specific lexical process. The results of the present study add to this literature by qualifying how this practice-driven optimization influences many of the lexical and semantic dimensions that are the focus of contemporary research into visual word recognition.

CONTROL VARIABLES

One present finding of interest is the observation of significant decreases in the contributions of word frequency with practice. We observed a clear facilitatory effect of frequency with increases in word frequency associated with faster RTs. Interestingly, in both the full (25,463 items) and restricted analysis (3723 items) the significant effects of frequency also formed a significant interaction with block, as the size of the frequency effect decreased with repeated practice.

There are numerous potential explanations for this diminishing utilization of frequency information by participants. In

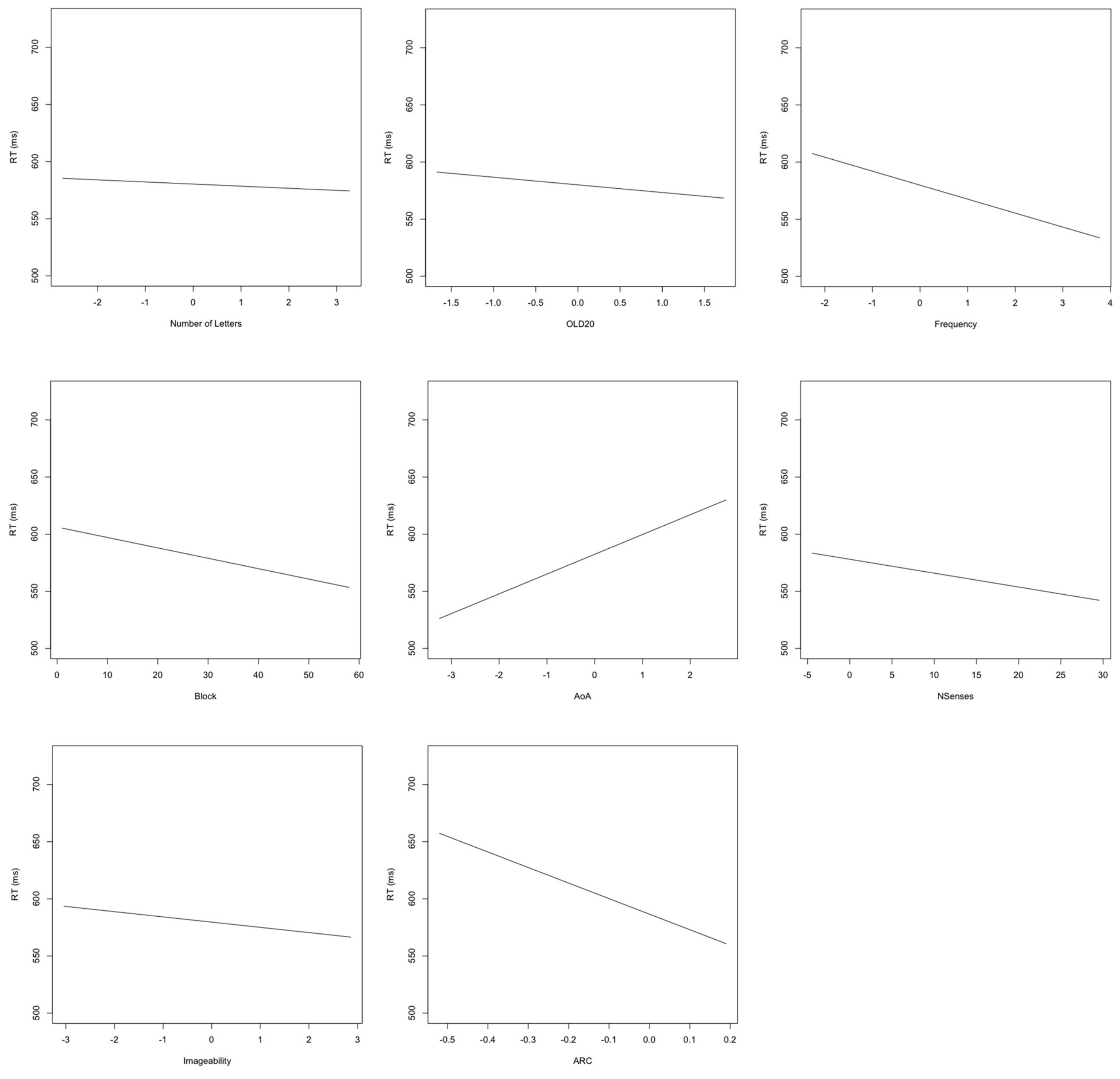


FIGURE 2 | Plots of fixed effects from a model fitted to LDT reaction times to 3723 words. Note: Length, number of letters (centered); OLD20, mean levenshtein distance to nearest 20 orthographic neighbors (centered); SUBTLEwf, log SUBTLEX frequency (centered); AoA, Age of acquisition (centered); NSenses, Number of Senses; IMG, Imageability (centered); ARC, semantic neighborhood density (centered).

diffusion model terms one could argue that faster information accumulation, coupled with narrower decision boundaries, could reduce the amount of time that frequency information has to accumulate, thereby reducing the influence of frequency on LDT RTs. However, one could just as easily argue that the decreasing effect of frequency may indicate that participants were, over time, adjusting to the large number of low-frequency words in the BLP stimulus set, rendering frequency-information less diagnostic of a decision. Both interpretations of the relationship between practice and frequency are admittedly *post hoc*, however, they serve to emphasize that these variables can conceivably influence both lexical and post-lexical processing.

One control variable that did not diminish in its contributions to LDT performance was AoA. Replicating numerous studies (Ellis and Lambon Ralph, 2000), we observed that words rated as being learned earlier in life were associated with faster responses in the LDT. Importantly, the AoA effect was highly consistent, and did not diminish with repeated practice. Though, correlated with word frequency, many researchers have suggested that the effect of AoA reflects a unique contribution that is

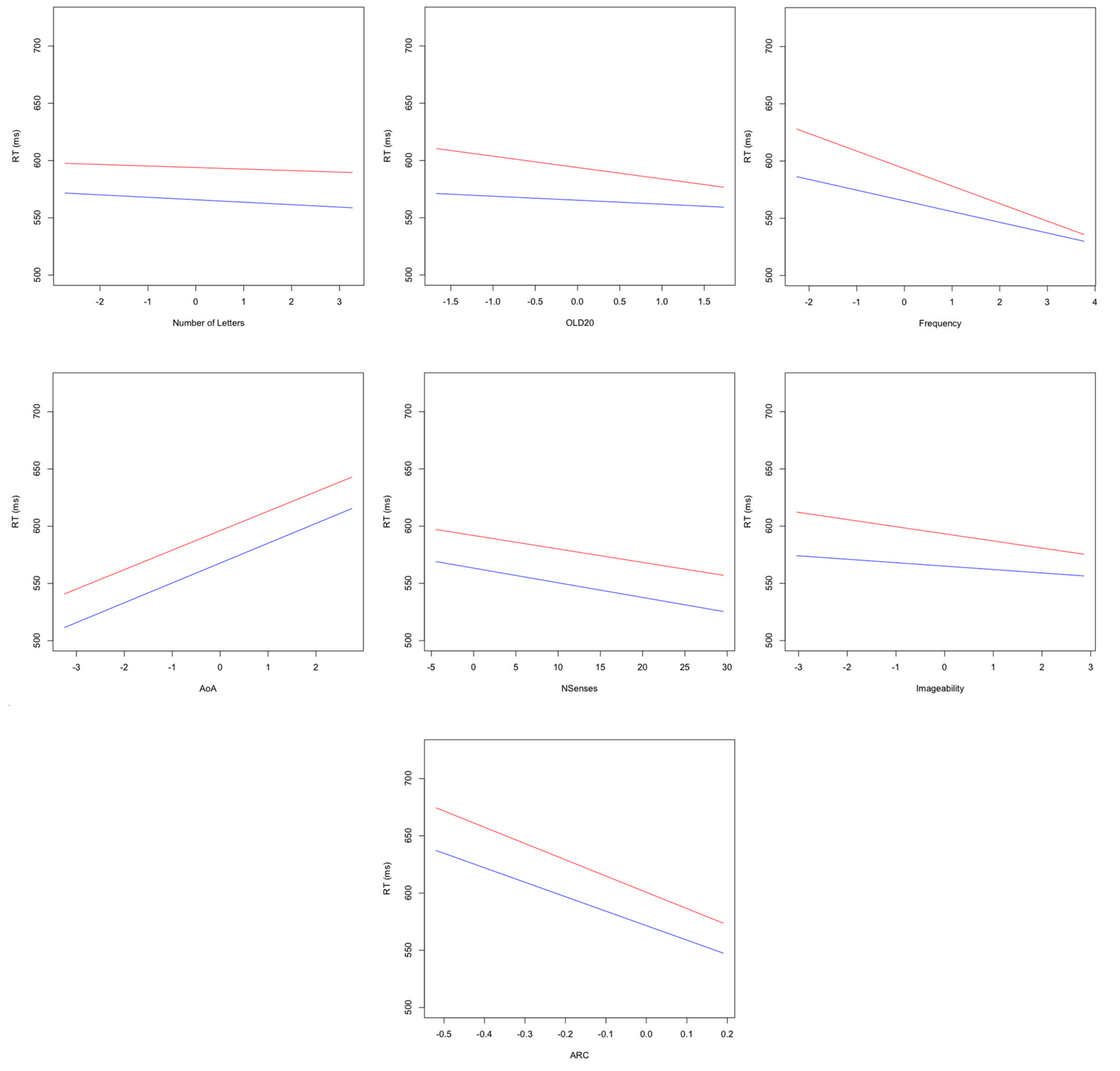


FIGURE 3 | Plots of fixed effects from a model fitted to LDT reaction times to 3723 words for early (less than block 28; red) and late (greater than block 28 but excluding the shortened block 57; blue) blocks. Note: Length, number of letters (centered); OLD20, mean levenshtein distance to nearest 20 orthographic neighbors (centered); SUBTLEwf, log SUBTLEX frequency (centered); AoA, Age of acquisition (centered); NSenses, Number of Senses; IMG, Imageability (centered); ARC, semantic neighborhood density (centered).

independent of that of word frequency (Juhasz, 2005; c.f., Zevin and Seidenberg, 2002). Whether AoA effects result from the influence of unmeasured cumulative word frequency (Zevin and Seidenberg, 2002) or have a semantic locus, as suggested by network models of AoA effects (Ellis and Lambon Ralph, 2000; Steyvers and Tenenbaum, 2005), the contributions of AoA are in some cases greater than the contributions of word frequency alone (Juhasz, 2005). The current findings also suggest a clear dissociation between frequency and AoA. As **Table 4** shows, there was a 39% reduction in the effect of word frequency between early and late blocks ($b = -15.30$ and -9.36 respectively). In contrast, AoA effects were indifferent to repeated practice, showing no appreciable change between early and late blocks ($b =$ 17.01 and 17.33 respectively), despite the overall increase in the speed of participants' LDT responses that accompanied practice. The current findings reveal that an analysis of repeated practice in the LDT might provide important insight into dissociating variables that are highly correlated, and demonstrate that the influence of AoA on LDT response times is pervasive and consistent.

SEMANTIC RICHNESS VARIABLES

Replicating findings from several previous studies we observed facilitatory effects for all of our semantic richness variables (Yap et al., 2012). However, while the individual contributions of these variables were consistently facilitatory, each variable responded uniquely to extensive practice in the LDT. In both the full and restricted analyses, ARC emerged as a significant predictor of reaction time in the LDT. This effect was qualified by a significant interaction with practice; however, though ARC effects decreased as participants gained more experience in the LDT the absolute decrease in the size of ARC effects was relatively small. As shown in **Tables 2** and **4**, between early and late blocks we observed a 6% decrease in the ARC effect for 25,463 items, and a 11% decrease for 3723 items. This stands in contrast to the effect of imageability which, like ARC, reached statistical significance and formed a significant interaction with block. Unlike ARC, the practice-driven decrease in imageability effects was relatively large, with a 52% decrease in the size of imageability effects between early and late blocks. This potential for practice effects to have strikingly disparate consequences for different measures of semantic richness is further highlighted by the finding that facilitatory effects of number of senses do not interact with practice. Indeed, as shown in **Table 4**, the fixed effects estimates for the number of senses are highly similar between early and late blocks ($b = -1.17$ and -1.28, respectively).

One theme to emerge from studies of semantic representation is that it is often useful to organize semantic dimensions into those that reflect object-based properties (i.e., semantic properties reflecting our immediate sensory experience with real-world exemplars of concepts) and those that reflect language-based properties (i.e., semantic properties reflecting our experiences processing the hierarchical statistical regularities that govern word-to-word usage in natural language; Buchanan et al., 2001). These divisions reflect distinct pathways by which humans can come to acquire and represent knowledge, and a diverse set of evidence suggests that both language- and object-based information can contribute during reading (Paivio, 1971; Buchanan et al., 2001; Solomon and Barsalou, 2004; Pulvermüller, 2010). Interestingly, recent evidence suggests that when task demands favor the shallow processing of meaning (e.g., as in the LDT; Lupker and Pexman, 2010) the language-based system reaches peak activation earlier, and contributes to the decision before the object-based simulation system (Simmons et al., 2008; Louwerse and Connell, 2011). One piece of evidence for this distinction comes from a study by Barsalou and colleagues, who asked participants to list properties of a provided word (Simmons et al., 2008; Santos et al., 2011). They found that the earliest listed associates were linguistically related to the cue, reflecting associative (e.g., *bee* -> *hive*) or phonological relationships (e.g., *self* -> *selfish*). Later associates tended to reflect properties that could emerge from situated simulation, such as properties of the environment (*golf* -> *sunshine*), or physical properties of the objects (*bee* -> *wings*). Subsequent analysis of fMRI data collected while participants were generating associates revealed that early (the first 7.5 s) property generation was moderated by classic language areas (e.g., Broca's area) while later generation (7.5–15 s) involved areas associated with mental imagery and episodic memory. The authors concluded that both language-based and object-based simulation systems contribute to the relatively shallow processing in the property generation task, however, the language-based system makes earlier contributions than the object based-system.

In the LDT, semantic processing is relatively shallow, and is thought to contribute to participants' decisions mostly in terms of feedback to orthography (Pexman et al., 2002). As participants gain experience with the LDT, one expectation is that practice-driven optimization will reduce the relative contributions of feedback, allowing participants to make their decisions while engaging in shallower semantic processing. This expectation is supported by the finding that competitive Scrabble players, who perform the LDT significantly faster than age-matched controls, seem to de-emphasize the role of meaning in their decision as evidenced by a significant reduction in the size of concreteness effects (Hargreaves et al., 2012). If practice driven efficiency leads to shallower semantic processing, those semantic processes that are faster will continue to contribute to the LDT decision (e.g., the language-based processes that reflect our histories of reading words) whereas the contributions of those systems that rely on deeper, situated simulation (e.g., the contributions of imagery information) will be disproportionately disrupted. One interesting feature of the current findings is that the observed dissociations in the effects of practice on the different descriptions of semantic richness also divide themselves between object-based (i.e., imagery) and language-based (i.e., ARC and number of senses) descriptions. The potential for repeated practice to encourage shallower semantic processing may account for the current data, in which practice-driven increases in LDT efficiency are associated with a reduction in participants' reliance on object-based (i.e., imagery) semantic information but this practice does not modulate their reliance on language-based (i.e., ARC and the number of senses) semantic information.

This interpretation, though admittedly *post-hoc*, connects with extant theories of the relative roles of linguistic and embodied information in informing reading, the unique characteristics of LDT expertise found among competitive Scrabble experts, and is also supported by a growing literature that utilizes multiple tasks in order to dissociate the effects of different descriptions of semantic richness (Yap et al., 2012). Numerous studies have highlighted the potential for word meaning to contribute to participants' judgments in the LDT. For years, the finding that words rated as being higher in mental imagery were responded to faster was taken as crucial evidence for the role of imagery-information in lexical processing (Paivio, 1971). The list of semantic dimensions continues to expand. For example, some researchers have taken advantage of surveys of dictionary definitions to characterize variability in word usage (e.g., number of senses; Yap et al., 2012), or have taken advantage of advances in computing in order to derive a variable that characterizes the history of a words' usage in a text-based corpus (e.g., ARC; Shaoul and Westbury, 2010). Like imageability, the contributions of these semantic effects to LDT performance are taken as evidence that these semantic dimensions shape lexical processing. There is mounting evidence that this diverse set of semantic richness variables each account for some unique aspect of meaning, and do not reflect the solitary contribution of a single underlying semantic factor. Utilizing cross-task comparisons, researchers have demonstrated that different descriptions of meaning play a unique role in performance that varies depending upon the demands of the task (Pexman et al., 2008; Yap et al., 2012). The current results add to this effort by revealing that, like task demands, changes in processing that occur with practice-driven optimization in the LDT reveal clear dissociations in the relative contributions made by different characterizations of semantic richness. This observation is supported by a pronounced diversity in the influence of practice on semantic richness effects, with some semantic variables showing large effects of practice (e.g., imageability), others showing more moderate effects (e.g., ARC), and still others showing no effect of practice at all (e.g., number of senses).

Our study provides the first investigation of the effects of repeated practice in the LDT on semantic richness effects. We selected predictor variables based on theoretical importance and the availability of a complete data set in order to maximize the sensitivity of what is an immense within-subjects study. We were able to find a complete set of values for 3723 words, and for a smaller set of predictors, 25,463 out of 28,730 words. This is a substantial improvement over previous investigations of semantic richness, and highlights the power of the BLP dataset. The observation of practice effects has implications for models of visual word recognition, particularly those that utilize only a single dimension of lexicality (e.g., MROM; Grainger and Jacobs, 1996). These models would have a difficult time explaining the present results precisely because we observed a dynamic tradeoff across multiple dimensions of the information utilized by participants in order to maximize efficiency. In order to explain the present findings, models of visual word recognition would need to incorporate the possibility that multiple dimensions of information can be emphasized or de-emphasized as a function of task demands, perhaps in a manner similar to that described in the attentional sensitization model (Kiefer and Martens, 2010).

In summary, the current study reveals that different dimensions of lexical and semantic information can display considerable variability in their utilization by participants over repeated practice. While some dimensions continue to provide information that is consistently diagnostic of a word decision (e.g., the Number of Senses and AoA) other dimensions become less important as the participant gains more familiarity with the demands of the decision, and with the kinds of items in the LDT. This suggests that the contributions of lexical and semantic information towards a lexical decision are dynamic. Though, the current results may be a function of the specific demands created by the LDT, they are consistent with a literature on practice effects that finds the influence of practice to play out across numerous dimensions, even for very basic tasks (Dutilh et al., 2009).

REFERENCES

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. New York, NY: Cambridge University Press.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412.

Bates, D. M. (2007). *Linear Mixed Model Implementation in Lme4*. Unpublished manuscript, Department of Statistics, University of Wisconsin-Madison, Madison, WI.

Balota, D. A., and Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 340–357.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. J. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. Gen.* 133, 283–316.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459.

Buchanan, L., Westbury, C., and Burgess, C. (2001). Characterizing semantic space: neighborhood effects in word recognition. *Psychon. Bull. Rev.* 8, 531–544.

Burgess, C. (1988). From simple associations to the building blocks of language: modeling meaning in memory with the HAL model. *Behav. Res. Methods Instrum. Comput.* 30, 188–198.

Burgess, C., and Lund, K. (2000). "The dynamics of meaning in memory," in *Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines*, eds E. Dietrich, and A. B. Markman (Mahwah, NJ: Erlbaum), 117–156.

Brysbaert, M., and New, B. (2009). Moving beyond Kuėera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American english. *Behav. Res. Methods* 41, 977–990.

Cortese, M. J., and Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behav. Res. Methods* 36, 384–387.

Cortese, M. J., and Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behav. Res. Methods* 40, 791–794.

Ellis, A. W., and Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: insights from connectionist networks. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1103–1123.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., and Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychon. Bull. Rev.* 16, 1026–1036.

Grainger, J., and Jacobs, M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychol. Rev.* 103, 518–565.

Halpern, D. F., and Wai, J. (2007). The world of competitive scrabble: novice and expert differences in visuospatial and verbal abilities. *J. Exp. Psychol. App.* 13, 79–94.

Hargreaves, I. S., Pexman, P. M., Zdrazilova, L., and Sargious, P. (2012). How a hobby can shape cognition: visual word recognition in competitive Scrabble players. *Mem. Cognit.* 40, 1–7.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture processing. *Psychol. Bull.* 131, 684–712.

Keuleers, E., and Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behav. Res. Methods* 42, 627–633.

Keuleers, E., and Brysbaert, M. (2011). Detecting inherent bias in lexical decision experiments with the LD1NN algorithm. *Ment. Lexicon* 6, 35–52.

Keuleers, E., Diependaele, K., and Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: a lexical decision study on 14,000 Dutch mono- and disyllabic words and non-words. *Front. Psychol.* 1:1–15. doi: 10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., and Brysbaert, M. (2011). The British lexicon project: lexical decision data for 28, 730 monosyllabic and disyllabic English words. *Behav. Res. Methods* 44, 287–304.

Kiefer, M., and Martens, U. (2010). Attentional sensitization of unconscious cognition: task sets modulate subsequent masked semantic priming. *J. Exp. Psychol. Gen.* 139, 464–489.

Louwerse, M., and Connell, L. (2011). A taste of words: linguistic context and perceptual simulation predict the modality of words. *Cognit. Sci.* 35, 381–398.

Lupker, S. J., and Pexman, P. M. (2010). Making things difficult in lexical decision: the impact of pseudo-homophones and transposed-letter nonwords on frequency and semantic priming effects. *J. Exp. Psychol. Lear. Mem. Cogn.* 36, 1267–1289.

New, B., Ferrand, L., Pallier, C., and Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: new evidence from the English lexicon project. *Psychon. Bull. Rev.* 13, 45–52.

Paivio, A., (1971). *Imagery and Verbal Processes*. New York, NY: Holt, Rinehart, and Winston.

Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number of features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.

Pexman, P., Hargreaves, I., Siakaluk, P., Bodner, G., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on visual word recognition. *Psychon. Bull. Rev.* 15, 161–167.

Pulvermüller, F. (2010). Brain-language research: where is the progress? *Biolinguistics* 4, 255–288.

Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20, 873–922.

Ratcliff, R., Thapar, A., and McKoon, G. (2006). Aging, practice, and perceptual tasks: a diffusion model analysis. *Psychol. Aging* 21, 353–371.

R Development Core Team. (2010). *R: A Language and Environment for Statistical Computing, Reference Index Version 2.14.2*. Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org

Santos, A., Chaigneau, S., Simmons, W. K., and Barsalou, L. (2011). Property generation reflects word association and situated simulation. *Lang Cogn.* 3, 83–119.

Shaoul, C., and Westbury, C. (2006). Word frequency effects in high-dimensional co-occurrence models: a new approach. *Behav. Res. Methods* 38, 190–195.

Shaoul, C., and Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behav. Res. Methods* 42, 393–413.

Simmons, W. K., Hamann, S., Harenski, C., Hu, X., and Barsalou, L. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *J. Physiol. (Paris)* 102, 106–119.

Solomon, K. O., Barsalou, L. W. (2004). Perceptual simulation in property verification. *Mem. Cognit.* 32, 244–259.

Stadthagen-Gonzalez, H., and Davis, C. J. (2006) The Bristol norms for age of acquisition, imageability and familiarity. *Behav. Res. Methods* 68, 598–605.

Steyvers, M., and Tenenbaum, J. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth, *Cognit. Sci.* 29, 41–78.

Stone, G. O., and Van Orden, G. C. (1993). Strategic control of processing in word recognition. *J. Exp. Psychol. Hum. Percepti. Perform.* 19, 744–774.

Tremblay, A. (2011). *LMERConvenienceFunctions: A Suite of Functions to Back-fit Fixed Effects and Forward-fit Random Effects.* [Computer software manual]. Available online at: http://CRAN.Rproject.org/package=LMERConvenienceFunctions (R package version 2.15).

Tuffiash, M., Roring, R. W., and Ericsson, K. A. (2007). Expert performance in SCRABBLE: implications for the study of the structure and acquisition of complex skills. *J. Exp. Psychol. Appl.* 13, 124–134.

Yap, M. J., Balota, D. A., Sibley, D. E., and Ratcliff, R. (2011). Individual differences in visual word recognition: insights from the English lexicon project. *J. Exp. Psychol. Hum. Percepti. Perform.* 38, 53–79.

Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speed pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., and Huff, M. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:72. doi: 10.3389/fnhum.2012.00073

Yarkoni, T., Balota, D. A., and Yap, M. J. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychon. Bull. Rev.* 15, 971–979.

Zevin, J., and Seidenberg, M. (2002). Age of acquisition effects in reading and other tasks. *J. Mem. Lang.* 47, 1–29.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 January 2012; accepted: 23 July 2012; published online: 14 August 2012.

Citation: Hargreaves IS and Pexman PM (2012) Does richness lose its luster? Effects of extensive practice on semantic richness in visual word recognition. Front. Hum. Neurosci. ***6****:234. doi: 10.3389/fnhum.2012.00234*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 03 May 2012
doi: 10.3389/fnhum.2012.00113

The influence of print exposure on the body-object interaction effect in visual word recognition

Dana Hansen[1], Paul D. Siakaluk[1]* and Penny M. Pexman[2]

[1] Department of Psychology, University of Northern British Columbia, Prince George, BC, Canada
[2] Department of Psychology, University of Calgary, Calgary, AB, Canada

Edited by:
Melvin Yap, National University of Singapore, Singapore

Reviewed by:
Chi-Shing Tse, The Chinese University of Hong Kong, Hong Kong
Mark Yates, University of South Alabama, USA

***Correspondence:**
Paul D. Siakaluk, Department of Psychology, University of Northern British Columbia, Prince George, BC V2N 4Z9, Canada.
e-mail: siakaluk@unbc.ca

We examined the influence of print exposure on the body-object interaction (BOI) effect in visual word recognition. High print exposure readers and low print exposure readers either made semantic categorizations ("Is the word easily imageable?"; Experiment 1) or phonological lexical decisions ("Does the item sound like a real English word?"; Experiment 2). The results from Experiment 1 showed that there was a larger BOI effect for the low print exposure readers than for the high print exposure readers in semantic categorization, though an effect was observed for both print exposure groups. However, the results from Experiment 2 showed that the BOI effect was observed only for the high print exposure readers in phonological lexical decision. The results of the present study suggest that print exposure does influence the BOI effect, and that this influence varies as a function of task demands.

Keywords: embodied cognition, perceptual symbol systems, motor simulation, print exposure, lexical conceptual processing

The body-object interaction (BOI) variable measures perceptions of the ease with which a human body can physically interact with a word's referent (Siakaluk et al., 2008a). As such, high BOI words (e.g., *mask*) refer to objects with which a human body can easily interact, whereas low BOI words (e.g., *ship*) refer to objects with which a human body cannot easily interact. In recent research the effects of BOI have been examined in a variety of visual word and object recognition tasks.

Siakaluk et al. (2008a) examined the effects of BOI in a lexical decision task; they manipulated BOI while controlling for the effects of numerous confound variables known to influence visual word recognition performance including, importantly, imageability and concreteness (i.e., their high BOI words and low BOI words were equally imageable and concrete). They reported a facilitatory effect of BOI, such that high BOI words were responded to more rapidly than low BOI words. Since this initial study, facilitatory effects of BOI have been reported for lexical decision in two studies using much larger sets of monosyllabic words (Tillotson et al., 2008; Siakaluk et al., 2011, Experiment 3) and in two studies using large sets of multisyllabic words (Bennett et al., 2011; Yap et al., 2012b).

The effects of BOI have also been examined in tasks in which responses are based primarily on phonological processing. Siakaluk et al. (2008a) used a phonological lexical decision task in which words, pseudohomophones (e.g., *brane*), and pronounceable non-words (e.g., *frane*) were used and the decision category was, "Does the item sound like a real English word?" In this task, "yes" responses were made to the words and pseudohomophones, whereas "no" responses were made to the pronounceable non-words. In addition, the effects of BOI have been examined in word naming (Bennett et al., 2011; Yap et al., 2012b) and picture naming tasks (Bennett et al.). As was the case for lexical decision, a facilitatory BOI effect was reported for each of these tasks.

Lastly, the effects of BOI have been examined in the semantic categorization task. Siakaluk et al. (2008b) used three different versions of this task. In Experiments 1A and 1B, the same set of high BOI words, low BOI words, and less imageable word foils were used. For Experiment 1A the decision category was, "Does the word refer to something that is easily imageable?", whereas for Experiment 1B the decision category was, "Does the word refer to something that is not easily imageable?" Thus, in the first experiment the experimental items required a "yes" response, whereas in the second experiment they required a "no" response. A facilitatory BOI effect was reported for both experiments. In Experiment 2, they conducted what they called a semantic lexical decision task. In this task, the high BOI words, low BOI words, and less imageable word foils were intermixed with pseudohomophones. There were two decisions that were required: first to decide if the item was a word or not, and second, if the item was a word, to decide if it was easily imageable or not. Again, a facilitatory BOI effect was reported, and, interestingly, it was significantly larger than that observed in Experiment 1A (in which only one decision needed to be made). Since this initial study, a facilitatory BOI effect has been reported in semantic categorization tasks (using the same decision category as used in Siakaluk et al., 2008b) in which verbal responses were used (Wellsby et al., 2011) and in which multisyllabic words were used (Bennett et al., 2011; Yap et al., 2012b, who used a "Does the word refer to something that is concrete?" decision category).

More recently, Tousignant and Pexman (2012) examined the effects of BOI in four versions of a semantic categorization task. The same set of high BOI words, low BOI words, and action words were used in each version, but the framing of the decision category varied between versions. The four decision categories (and sets of instructions) were: "Is it an entity?" (press the left button for entities and the right button for non-entities); "Is it an entity or an action?" (press the left button for entities and the right button for actions); "Is it an action or an entity?" (press the left button for actions and the right button for entities); and "Is it an action?" (press the left button for actions and the right button for non-actions). Tousignant and Pexman reported a facilitatory BOI effect in all three versions in which the instructions included entity words as part of the decision category, but not in the version in which the instructions did not include entity words as part of the decision category. They proposed that BOI information is used under conditions in which object information is made salient, such as when participants are expecting to see entity words (which are concrete nouns that refer to concepts with which human bodies can physically interact), but not for action words (which are verbs that refer to concepts with which human bodies cannot physically interact).

There are two frameworks that, when combined, have been used to provide an explanatory account for facilitatory effects of BOI. The first is an influential embodied cognition framework known as perceptual symbol systems (Barsalou, 1999, 2008). Embodied cognition more generally is the theoretical perspective that much of human cognition is acquired through (or grounded in) sensorimotor experience with the environment (Clark, 1997; Lakoff and Johnson, 1999; Wilson, 2002; Pecher and Zwaan, 2005). There are two key assumptions of the perceptual symbol systems framework that are relevant for an explanatory account of facilitatory BOI effects. The first assumption is that lexical conceptual knowledge is multimodal. That is, there are multiple neural systems involved in the acquisition and retrieval of lexical conceptual knowledge. Among these are neural systems dedicated to processing sensory knowledge (e.g., visual, auditory), emotional knowledge (e.g., fear, excitement), introspective knowledge (e.g., association, thought), and, most relevant in accounting for the facilitatory effects of BOI, motor, kinesthetic, and proprioceptive knowledge (e.g., physically interacting with objects, internal feedback from muscles and joints). The second assumption is that retrieving lexical conceptual knowledge from memory involves the process of simulation. Simulation refers to the partial re-enactment of the states of the various neural systems that were involved at the time of encoding. Importantly, these assumptions were supported by a recent fMRI study examining the effects of BOI in a semantic categorization task that used the imageability decision category (Hargreaves et al., 2012a). Hargreaves and colleagues reported that in addition to a large facilitatory behavioral effect of BOI, greater activation was observed for high BOI words than for low BOI words in the left inferior parietal lobule (i.e., the supramarginal gyrus, BA 40), which is a brain region associated with kinesthetic memory (Grèzes and Decety, 2001; Péran et al., 2010).

We have previously extended the perceptual symbol systems framework to provide an explanatory account for facilitatory effects of BOI in the following way (e.g., Siakaluk et al., 2008a,b). As noted, high BOI words refer to objects that human bodies can easily physically interact with, whereas low BOI words refer to objects that human bodies cannot easily physically interact with. Thus, the former types of words will develop and eventually elicit richer motor, kinesthetic, and proprioceptive representations than will the latter types of words. Stated another way, using the terminology of perceptual symbol systems, high BOI words will develop and eventually elicit richer motor, kinesthetic, and proprioceptive simulations (we will hereafter simply refer to these different types of simulations as motor simulations) than will low BOI words.

The second framework that is relevant to an explanatory account for facilitatory effects of BOI is the semantic feedback framework (Hino and Lupker, 1996; Pexman and Lupker, 1999; Pecher, 2001; Hino et al., 2002). This framework has three important assumptions and two mechanisms by which facilitatory effects of BOI may arise. The first assumption is that different word characteristics are processed in different, dedicated sets of units. That is, orthographic knowledge is processed within orthographic units, phonological knowledge is processed within phonological units, and semantic knowledge is processed within semantic units. The second assumption is that these three sets of units are interconnected such that the processing of one set of units may influence the processing of another set of units. The third assumption is that the impact of the processing of one set of units on another set of units is dependent on the nature of the connections between the two sets of units. One mechanism is involved in tasks in which responses are based primarily on semantic processing (e.g., semantic categorization). A facilitatory BOI effect arises in these tasks because high BOI words elicit richer semantic activation (i.e., richer motor simulations) within the semantic units that leads to faster settling on a semantic representation and hence faster semantic categorization latencies. The other mechanism is involved in tasks in which responses are based primarily on either orthographic processing (e.g., lexical decision) or phonological processing (e.g., phonological lexical decision). A facilitatory BOI effect arises in these tasks because high BOI words elicit greater semantic activation (i.e., richer motor simulations) within the semantic units, which then sends stronger semantic feedback to the orthographic units and to the phonological units, leading to faster settling on an orthographic representation or phonological representation and hence faster lexical decision latencies and phonological lexical decision latencies, respectively. Thus, facilitatory BOI effects have been explained by this combination of the perceptual symbol systems framework for embodied semantic knowledge and the semantic feedback framework for the visual word recognition system.

Zwaan (2008) suggested several avenues of future research for those interested in studying embodiment effects in language, one of which was "to examine more closely the role of prior experience . . . in language comprehension" (p. 172). Although perhaps not exactly what Zwaan had in mind when he gave this recommendation, we were interested in a related idea: examining whether prior *reading* experience (i.e., print exposure) would modulate the facilitatory effects of BOI in visual word recognition, using semantic categorization and phonological lexical decision tasks. Before presenting in more detail the purpose of the present study, we will discuss the lexical integrity hypothesis

(Yap et al., 2009), which is integral to the predictions we make below regarding the influence of print exposure on facilitatory effects of BOI.

Yap et al. (2009) developed the lexical integrity hypothesis (see also the lexical quality hypothesis, e.g., Perfetti, 1992; Perfetti and Hart, 2002; Andrews and Bond, 2009) to account for several findings regarding the influence of semantics (and lexical variables such as print frequency) in the literature. For example, in their study, Yap et al. examined the joint influence of semantic priming and word frequency in lexical decision. Their primary findings were that these two variables have interactive effects (i.e., larger effects of semantic priming for low-frequency words than for high-frequency words) for readers with less vocabulary knowledge, but have additive effects (i.e., similar effects of semantic priming for low-frequency words and for high-frequency words) for readers with more vocabulary knowledge. Yap et al.'s notion of lexical integrity accounts for these findings in the following way. Readers with more vocabulary knowledge develop higher integrity orthographic representations that are closer to recognition threshold, whereas readers with less vocabulary knowledge have lower integrity orthographic representations that are further removed from recognition threshold. Importantly, the difference in lexical integrity between the two types of readers is likely to be larger for lower frequency words. Yap et al. state it this way, "a medium-frequency word for a high-lexical-integrity individual is likely to be a low-frequency word for a low-lexical-integrity individual" (p. 306). In general, because orthographic representations further away from recognition threshold require more lexical conceptual processing before lexical decisions can be made, they would benefit more from, say, semantic priming. Thus, Yap et al. predicted that, "one might actually expect individuals with *lower* integrity representations to show a larger influence of semantic context than those with higher integrity representations" (p. 306; emphasis in original). This reasoning can also account for the recent demonstration that non-expert Scrabble players showed larger effects of concreteness in lexical decision than did competitive Scrabble players, who have considerable lexical knowledge (Hargreaves et al., 2012b).

We propose to extend the lexical integrity hypothesis in the following ways. First, we assume that readers with more print exposure, in addition to developing higher integrity orthographic representations, also develop higher integrity semantic representations and higher integrity phonological representations. Second, as noted above in our discussion of the semantic feedback framework, the recognition threshold that needs to be exceeded for responding depends on task demands. That is, semantic categorizations are based primarily on semantic processing, and responses are made available when semantic representations exceed recognition threshold; lexical decisions are based primarily on orthographic processing, and responses are made available when orthographic representations exceed recognition threshold; and phonological lexical decisions are based primarily on phonological processing, and responses are made available when phonological representations exceed recognition threshold. Recall that, according to the semantic feedback framework, in the latter two cases, orthographic processing and phonological processing may be influenced by semantic feedback.

THE PRESENT STUDY

The purpose of the present study was to examine the influence of print exposure on facilitatory effects of BOI in semantic categorization and phonological lexical decision. To measure print exposure we used the Canadian version of the Author Recognition Test (ART) (Chateau and Jared, 2000). This version is based on the ART originally developed by Stanovich and West (1989). Stanovich and West developed the ART to overcome social-desirability effects in the assessment of print exposure. The ART consists of a list of names, some of which are popular writers of books, magazine articles, and/or newspaper columns (e.g., Margaret Atwood) and some of which are not (e.g., Anne Cunningham). Participants are instructed to only put a check mark next to the names of the individuals whom they know to be writers, and guessing is discouraged because incorrect responses are penalized. The ART has received extensive validation, such that ART scores are associated with early reading ability (Cunningham and Stanovich, 1997), reading experience (Cunningham and Stanovich, 1990), and, most importantly for the present study, vocabulary knowledge (West et al., 1993; Lee et al., 1997), such that higher ART scores predict greater vocabulary knowledge. Thus, the benefits of using the ART in measuring print exposure are that it avoids concerns of social-desirability effects, it is reliably associated with many characteristics of reading experience and ability, and it is relatively cheap to use in terms of required resources (e.g., it is freely available, takes only a few minutes for participants to complete, and is easy to score).

As noted, the present research is concerned with examining the effects of print exposure on facilitatory effects of BOI in semantic categorization and phonological lexical decision. We will first address the experimental procedure, and the predictions and results of the influence of print exposure on BOI in semantic categorization. We will postpone addressing these issues regarding the phonological lexical decision task until after our discussion of the semantic categorization task.

As described below, a high print exposure group and a low print exposure group performed a semantic categorization task in which the decision category was, "Does the word refer to something that is easily imageable?" We used a go/no-go procedure (in which participants responded only to the experimental items), rather than a yes/no procedure (in which participants would respond to both the experimental items and the foil items). Siakaluk et al. (2003) proposed that the go/no-go procedure should elicit more extensive semantic processing than the yes/no procedure using the following reasoning: because overt responses under go/no-go conditions are made only to the experimental items, this may lead participants to adopt a stricter decision criterion to ensure correct responses are made, which would allow for more extensive semantic processing to occur. Further, Siakaluk et al. (2003) predicted that there should be longer response latencies and lower error rates using the go/no-go procedure than the yes/no procedure. These two predictions were supported in their study, as well as in Siakaluk et al. (2007). Most importantly, in both these studies, semantic richness effects were more robust using the go/no-go procedure.

Based upon the lexical integrity hypothesis (Yap et al., 2009), we made the following two predictions. First, there should be a main effect of print exposure such that high print exposure readers should exhibit faster response latencies than low print exposure readers. Second, the facilitatory BOI effect should be smaller for high print exposure readers than for low print exposure readers. We made these two predictions based on the reasoning that: (1) high print exposure readers should have higher integrity semantic representations that should be closer to recognition threshold and should thus benefit less from the richer motor simulations evoked by high BOI words; whereas (2) low print exposure readers should have lower integrity semantic representations that are further from recognition threshold and should thus benefit more from the richer motor simulations evoked by high BOI words.

EXPERIMENT 1

METHODS

Participants

Ninety-two undergraduate students from the University of Northern British Columbia participated in the experiment for bonus course credit. All participants were native English-speakers and reported normal or corrected-to-normal vision.

Participants were administered a Canadian version of the ART (Chateau and Jared, 2000) after they completed the semantic categorization task (described below)[1]. For the data analyses, two groups of participants were created using a quartile split of the ART scores. As such, 23 participants were assigned to the high print exposure group (with a mean ART score of 17.0 and a range of 11–33) and 23 participants were assigned to the low print exposure group (with a mean ART score of 3.9 and a range of 3–5).

Stimuli

The experimental stimuli consisted of the 24 high BOI words (e.g., *mask*) and the 24 low BOI words (e.g., *ship*) used in Siakaluk et al. (2008b). The two sets of words were matched for print length, objective print frequency (using HAL log frequency norms from the English Lexicon Project database; Balota et al., 2007), subjective frequency, orthographic and phonological neighborhood size, phonological feedback inconsistency, contextual dispersion, semantic distance, number of features, senses, and associates, and importantly, concreteness and imageability (all p's > 0.15). In addition, each word only had one entry in the ITP Nelson Canadian Dictionary, (1997) and all had noun definitions listed first. The descriptive statistics for the experimental stimuli are listed in **Table 1**. The 48 less imageable noun foils (e.g., *fate*) used in Siakaluk et al. (2008b) were also used, and they had a mean imageability rating of 2.6 and a mean printed frequency of 18.9. All the stimuli are listed in the **Appendix**.

[1]The version of the Canadian ART we administered had 58 names of authors and 57 names of non-authors, for a total of 115 names.

Apparatus and procedure

The stimuli were presented on a color VGA monitor driven by a Pentium-class microcomputer running DirectRT software (http://www.empirisoft.com/DirectRT.aspx). A trial was initiated by a fixation marker appearing in the center of the computer display. The fixation marker was presented for 1 s and was then replaced by a word. The participants' task was to decide whether the words were easily imageable or not. Participants were instructed to press the "?" key on the computer keyboard when the imageable words were presented, and to make no response when the less imageable words were presented. For trials in which no response was made, stimulus items remained on the computer display for 2.5 s, and were then removed and replaced by the fixation marker. Participants were further instructed to make their responses as quickly but as accurately as possible. Response latencies were measured to the nearest ms. The order in which the stimuli were presented was separately randomized for each participant. The intertrial interval was 2 s.

Before beginning the experiment, each participant completed 20 practice trials that consisted of 10 imageable words and 10 less imageable words. All the practice stimuli were similar in printed frequency to the stimuli used in the experiment.

RESULTS

Data for the low BOI word *tribe* were excluded from the analyses because the error rate for this item was 52.2%. The removal of this item did not affect the matching for the two sets of BOI words for any of the control variables listed above (all p's > 0.20).

Outliers were identified in the following manner. First, response latencies faster than 250 ms or slower than 2000 ms were considered outliers. Second, for each participant, response latencies greater than 2.5 *SD*s from the cell mean of each condition were considered outliers. Using this procedure, a total of 69 observations (3.2% of the data) were removed from the data-set.

Response latencies for correct responses and error percentages were analyzed using a 2 (print exposure: high, low) × 2 (BOI: high, low) mixed-model analysis of variance (ANOVA). Both subject (F_1) and item (F_2) analyses were conducted. In the subject analyses, print exposure was a between-subjects variable and BOI was a within-subjects variable. In the item analyses, print exposure was a within-items variable and BOI was a

Table 1 | Mean characteristics for word stimuli.

Word type	BOI	Plen	OFreq	SFreq	N	PN	PFI	CD	SemD	NumF	NumS	NumA	Conc	Image
High BOI	5.3	4.5	17.3	3.5	7.1	14.7	3.0	0.7	307.7	3.4	5.7	14.3	5.9	6.3
Low BOI	3.3	4.4	17.9	3.6	6.4	13.3	3.1	0.7	307.3	3.6	4.7	13.7	5.8	6.3

Note: BOI, body-object interaction; Plen, print length; OFreq, objective frequency; SFreq, subjective frequency; N, orthographic neighborhood size; PN, phonological neighborhood size; PFI, phonological feedback inconsistency; CD, contextual dispersion; SemD, semantic distance; NumF, number of features; NumS, number of senses; NumA, number of associates; Conc, concreteness; Image, imageability.

between-items variable. Unless noted, all effects were statistically significant at $p < 0.05$. The mean response latencies for correct responses and mean error percentages for the experimental words are presented in **Table 2**. The error percentages for the less imageable word foils were 8.5% for the high print exposure group and 7.3% for the low print exposure group.

Response latency analysis

For the response latency data, there was an effect of print exposure, $F_1(1, 44) = 5.05$, MSE = 34,743.24, $\eta^2 = 0.10$; $F_2(1, 45) = 110.10$, MSE = 1593.21, $\eta^2 = 0.71$, with the high print exposure group responding on average 88 ms faster than the low print exposure group. There was an effect of BOI, $F_1(1, 44) = 84.15$, MSE = 1141.79, $\eta^2 = 0.66$; $F_2(1, 45) = 9.14$, MSE = 11,532.60, $\eta^2 = 0.17$, with responses to high BOI words on average 64 ms faster than responses to low BOI words. There was also an interaction between print exposure and BOI, $F_1(1, 44) = 4.33$, MSE = 1141.79, $\eta^2 = 0.09$; $F_2(1, 45) = 4.96$, MSE = 1593.21, $\eta^2 = 0.10$. This interaction was followed up by analyzing the effects of BOI for each print exposure group separately. For the high print exposure group, the 50 ms BOI effect was significant, $t_1(22) = 5.03$, SEM = 9.93, $\eta^2 = 0.54$; $t_2(45) = 2.76$, SEM = 17.62, $\eta^2 = 0.15$. For the low print exposure group, the 79 ms BOI effect was significant, $t_1(22) = 7.93$, SEM = 9.99, $\eta^2 = 0.74$; $t_2(45) = 3.00$, SEM = 28.41, $\eta^2 = 0.17$. The interaction indicates that although facilitatory effects of BOI were observed for both print exposure groups, the effect was significantly larger for the low print exposure group than for the high print exposure group.

Error analysis

For the error data, there was no effect of print exposure, both Fs < 1, but there was an effect of BOI, $F_1(1, 44) = 13.27$, MSE = 12.64, $\eta^2 = 0.23$; $F_2(1, 45) = 4.97$, MSE = 32.40, $\eta^2 = 0.10$, with responses to high BOI words on average 2.7% more accurate than responses to low BOI words. There was no interaction between print exposure and BOI, $F_1(1, 44) = 2.03$, $p = 0.16$, MSE = 12.64; $F_2(1, 45) = 1.33$, $p = 0.25$, MSE = 15.34.

Table 2 | Mean raw response latencies (in ms) and standard errors, mean error percentages and standard errors, and mean transformed response latencies (in *Z*-scores) and standard errors for Experiment 1.

Word type	High print exposure		Low print exposure	
	M	**SE**	***M***	**SE**
RESPONSE LATENCIES				
High BOI	667	21.9	740	30.7
Low BOI	717	24.2	819	33.4
BOI effect	+50		+79	
RESPONSE ERRORS				
High BOI	0.6	0.3	1.4	0.6
Low BOI	4.3	1.2	3.0	1.0
BOI effect	+3.7		+1.6	
TRANSFORMED RESPONSE LATENCIES				
High BOI	−0.155	0.02	−0.194	0.02
Low BOI	0.170	0.03	0.202	0.02
BOI effect	0.325		0.396	

Note: BOI, body-object interaction.

DISCUSSION

The purpose of the present experiment was to examine the influence of print exposure on the facilitatory BOI effect in semantic categorization. More specifically, we predicted that there would be a main effect of print exposure (i.e., faster response latencies for the high print exposure readers than for the low print exposure readers) and that the facilitatory BOI effect would be smaller for high print exposure readers than for low print exposure readers. Both of these predictions were borne out by our results.

The above two findings can be accounted for in the following way. Based on the reasoning underlying the lexical integrity hypothesis (Yap et al., 2009), it is likely that the high print exposure readers developed higher integrity semantic representations than did the low print exposure readers. That is, words of a certain objective frequency (in the case of the present study the words were of low frequency) would be closer to semantic recognition threshold for the high print exposure readers due to their more frequent reading experience (to the extent that reading experience is measured by the ART), and would, therefore, result in both faster overall responding, and requiring less simulation of motoric knowledge before responding. The latter outcome would result in less benefit from motor simulation for the high print exposure readers and hence a reduced facilitatory BOI effect. In other words, there is less need for motor simulation to influence the settling of a semantic representation for high print exposure readers.

The above findings (i.e., a main effect of print exposure and a print exposure by BOI interaction) are consistent with the findings of the two studies we noted in our description above of the lexical integrity hypothesis. More specifically, Yap et al. (2009) reported the following two results. First, there was a main effect of group, such that participants from Washington University in St. Louis (WUSTL) had significantly faster overall lexical decision response latencies than did participants from University at Albany, State University of New York (SUNY-A), who also, interestingly, had lower levels of vocabulary knowledge. Second, there was a significant interaction between group, priming, and word frequency, such that the largest priming effect was observed for low-frequency words for the SUNY-A participants. Furthermore, Hargreaves et al. (2012b) reported that non-expert Scrabble players (who had less extensive lexical knowledge) had slower overall lexical decision latencies and larger effects of concreteness than competitive Scrabble players. Thus, our observed findings of a main effect of print exposure and a larger facilitatory effect of BOI for the low print exposure readers are consistent with previous findings in the literature, and with the notion that lexical integrity influences the effects of semantic richness in the semantic categorization task.

Despite the present finding that the facilitatory BOI effect was smaller for the high print exposure readers than for the low print exposure readers, it is worthwhile to emphasize that the effect was still quite large for the former type of reader. This suggests that even though it is likely that the low-frequency words used in the present study were associated with higher semantic recognition thresholds for the high print exposure readers, motor simulation still played an important role in lexical conceptual processing for readers with relatively more reading experience in the semantic categorization task, a task in which responses are based on semantic processing. In other words, knowledge gained through sensorimotor experience still exerted a facilitatory effect on lexical conceptual processing for readers who putatively have higher integrity semantic representations. This is an important discovery.

Some researchers, however, might raise an alternative interpretation of our findings (e.g., Faust et al., 1999; Yap et al., 2012a)[2]. For example, Faust et al. stated that "response latencies for different groups are often linearly related, leading to an increased likelihood of finding spurious overadditive interactions in which the slower group produces a larger treatment effect" (p. 777). According to this viewpoint, there are two factors that may contribute to observed group differences for the variable under examination. The first factor is *processing rate*, which is an individual differences factor, such that slower readers will have on average slower processing rates than faster readers. The second factor is *processing amount*, which refers to the information processing requirements of the task (e.g., in our Experiment 1, the accumulation of evidence that a word is easily imageable in order to make a "yes" response).

According to this viewpoint, the print exposure by BOI interaction observed in Experiment 1 may be explained in one of two ways. First, it may be that the larger facilitatory BOI effect for the low print exposure readers was primarily due to their receiving greater benefits of motor simulation for the high BOI words, attributable to the processing amount factor discussed above. Second, it may be that the larger facilitatory BOI effect for the low print exposure readers was simply due to their having slower overall response latencies, attributable to the processing rate factor. Needless to say, the second of these explanations is less theoretically interesting, due to the claim that observed group differences are merely due to a correlation with processing rate and not due to differential levels of motor simulation elicited by high BOI and low BOI words for the high print exposure and low print exposure readers.

To differentiate between these two possible explanations, Faust et al. (1999) suggested transforming raw response latencies to a common scale, which would factor out overall group differences in processing rate. By so doing, according to this viewpoint, if the group by variable interaction is still observed after response latency transformation, then it can be attributed to differences in processing amount between the two groups (which would be the theoretically more interesting outcome), but if the group by variable interaction is not observed after the transformation, then the original finding can be simply attributed to processing rate (which would be the theoretically less interesting outcome). Thus, we transformed the raw response latencies using the *z*-score transformation procedure. We then analyzed our *z*-score transformed response latency data using a 2 (print exposure: high, low) × 2 (BOI: high, low) mixed-model ANOVA. Because we believe this is a much more conservative test of our data, we employed planned comparisons examining the effect of BOI for each print exposure group separately, regardless of whether the print exposure by BOI interaction was significant. The mean transformed response latencies for correct responses for the experimental words are presented in **Table 2**.

[2]We thank Melvin Yap for bringing this to our attention.

Z-score response latency analysis

There was, of course, no effect of print exposure, $F_1(1, 44) = 1.50$, $MSE = 0.001$; $F_2 < 1$. There was an effect of BOI, $F_1(1, 44) = 119.77$, $MSE = 0.025$, $\eta^2 = 0.73$; $F_2(1, 45) = 8.94$, $MSE = 0.22$, $\eta^2 = 0.17$, with a *z*-score difference between the high BOI words and the low BOI words of 0.292 (and the high BOI words having faster latencies). There was no interaction between print exposure and BOI, both $F\text{s} < 1.40$. For the high print exposure group, the *z*-score difference between the high BOI words and the low BOI words was 0.325 and was significant, $t_1(22) = 6.42$, $SEM = 0.05$, $\eta^2 = 0.65$; $t_2(45) = 2.76$, $SEM = 0.09$, $\eta^2 = 0.15$. For the low print exposure group, the *z*-score difference between the high BOI words and the low BOI words was 0.396 and was significant, $t_1(22) = 9.40$, $SEM = 0.04$, $\eta^2 = 0.80$; $t_2(45) = 2.97$, $SEM = 0.11$, $\eta^2 = 0.16$. The analysis of the *z*-score transformed response latencies suggests that the original finding of a larger facilitatory effect of BOI for the low print exposure readers was largely due to differences of processing rate between the two print exposure groups, rather than an effect of differential impact of motor simulation elicited by the high BOI words and the low BOI words. We will address this issue in more detail in the General Discussion section.

EXPERIMENT 2

The experimental design used in Experiment 1 allowed us to examine the effect of print exposure on BOI in a task (semantic categorization) in which responses were based primarily on semantic processing. Recall that according to the semantic feedback framework, semantic processing *per se* is one of two mechanisms by which semantic richness effects may arise in the visual word recognition system. Further recall that the second such mechanism is feedback from semantics to orthography or to phonology, and thus semantic richness effects can be examined in tasks in which responses are based primarily on either orthographic processing (e.g., lexical decision) or phonological processing (e.g., phonological lexical decision).

As noted, ART scores are associated with various aspects of reading, such as early reading ability, reading experience, and vocabulary knowledge. ART scores are also associated with orthographic and phonological processing differences in skilled readers. For example, Chateau and Jared (2000) used the ART to create groups consisting of either high print exposure readers or low print exposure readers (the two print exposure groups were matched for performance on the comprehension subtest of the Nelson–Denny Reading Test). Chateau and Jared examined orthographic processing using both a homophone choice task, in which participants must choose the correct homophone for a given category (e.g., FRUIT: *pear-pair*), and a lexical decision task with pseudohomophones as foils; they examined phonological processing using a pseudoword naming task in which participants named aloud pseudowords that were not pseudohomophones (e.g., *shup*); and they jointly examined orthographic and phonological processing in a form priming task (e.g., *touch-couch*). In each task, the high print exposure readers exhibited more efficient orthographic and/or phonological processing such that: (1) they were significantly faster and more accurate in the homophone choice, lexical decision, and pseudoword naming tasks; (2) they had smaller effects of word frequency and orthographic typicality (i.e., more quickly accepted non-wordlike words and rejected more wordlike pseudohomophones) in the lexical decision task; and (3) they quickly activated orthographic representations when reading high-frequency words, but soon after more strongly activated phonological representations of those words in the form priming task (see Stanovich and West, 1989, for a similar set of results). Chateau and Jared concluded that the increased reading experience of the high print exposure readers led to development of more efficient orthographic and phonological processing skills as compared to the low print exposure readers.

Sears et al. (2008) also used the ART to create groups consisting of either high print exposure readers or low print exposure readers (but unlike Chateau and Jared, 2000, their two print exposure groups were not matched for performance on any other measure). Sears et al. examined the effect of print exposure on the effects of word frequency and neighborhood size (the finding that words with many orthographic neighbors are responded to more quickly than words with few orthographic neighbors; see Siakaluk et al., 2002) in two lexical decision tasks, one in which regular non-words were used and another in which pseudohomophones were used. They reported that high print exposure readers were significantly faster and more accurate, and had significantly smaller effects of word frequency and neighborhood size than did the low print exposure readers only in the lexical decision task in which pseudohomophones were used. In the lexical decision task in which regular non-words were used there were no effects of print exposure. Sears et al. interpreted their findings as indicating that low print exposure readers, "use phonological information to compensate for less efficient orthographic processing skills and this leads to larger word frequency and neighborhood size effects in a lexical decision task when phonology cannot be used to discriminate words from nonwords (such as when pseudohomophones are used in the task)" (p. 289).

Lastly, Unsworth and Pexman (2003) examined phonological processing differences between more skilled readers and less skilled readers (as measured by the ART, and the comprehension and vocabulary subtests of the Nelson–Denny Reading Test) in lexical decision and phonological lexical decision tasks. Their most robust findings were that the more skilled readers responded more quickly and showed no effects of spelling-to-sound regularity in either task, whereas the less skilled readers showed significant effects of spelling-to-sound regularity in both tasks. Unsworth and Pexman concluded that the more skilled readers had developed more efficient phonological processing skills as compared to the less skilled readers. In other words, the more skilled readers had developed more efficient orthographic-to-phonological mappings than the less skilled readers.

As noted, a second purpose of the present study was to examine the effects of print exposure on BOI in the phonological lexical decision task, which would allow us to determine whether print exposure influences the effects of BOI on semantic feedback to phonology. As the above review of the literature amply demonstrates, print exposure influences phonological (and orthographic) processing. Experiment 2 provided an opportunity to extend this work by examining whether print exposure influences orthographic-to-semantics-to-phonological mappings, because according to the semantic feedback framework, phonological lexical decision responses are based primarily on phonological processing, but can be modulated by feedback from semantics to phonology.

There are two possible outcomes of the influence of print exposure on BOI in the phonological lexical decision task. The first outcome, according to the lexical integrity hypothesis (Yap et al., 2009), is that a larger effect of BOI should be observed with the low print exposure readers, because their phonological representations should be further away from recognition threshold and would thus benefit more from motor simulation. The second outcome, according to the literature demonstrating that high print exposure readers have more efficient phonological processing, is that a larger effect of BOI may be observed with these readers, because the effects of semantic feedback, derived from motor simulation in semantics, should be more efficiently mapped onto phonological processing for this group of readers.

METHODS

Participants

Seventy-two undergraduate students from the University of Northern British Columbia participated in the experiment for bonus course credit. All participants were native English-speakers and reported normal or corrected-to-normal vision. None of these individuals participated in Experiment 1.

Participants were administered the ART after they completed the phonological lexical decision task (described below). For the data analyses, two groups of participants were created by taking the top and bottom 40% of the ART scores. As such, 29 participants were assigned to the high print exposure group (with a mean ART score of 17.0 and a range of 11–31) and 28 participants were assigned to the low print exposure group (with a mean ART score of 5.6 and a range of 2–9)[3].

Stimuli

The same set of experimental words used in Experiment 1, and the 48 pseudohomophones and 96 non-words used in the phonological lexical decision task by Siakaluk et al. (2008a) were used in the present experiment. Due to very low error rates, the low BOI item *tribe* was retained in the present set of analyses. The non-word and pseudohomophone stimuli are listed in the **Appendix**.

Apparatus and procedure

The apparatus and procedure were identical to Experiment 1 except for the following. The participants' task was to decide

[3]The mean ART score for the low print exposure group in Experiment 2 was higher (and had a larger range) than for the low print exposure group in Experiment 1. This was unavoidable due to the different distributions of ART scores for each group in the two experiments.

whether each letter string sounded like a real English word. Participants were instructed to press the "?" key on the computer keyboard when letter strings that sounded like real English words were presented (i.e., for the experimental words and the pseudohomophones), and to make no response when letter strings that did not sound like real English words were presented (i.e., for the non-words).

Before beginning the experiment, each participant first completed 20 practice trials that consisted of five words similar in normative frequency to the experimental items, five pseudohomophones, and 10 non-words.

RESULTS

Outliers were identified in the same fashion as in Experiment 1. A total of 102 observations (3.7% of the data) were removed from the data-set. Only six errors were made (two errors to the high BOI words and four errors to the low BOI words) across all 57 participants, thus no error analysis was conducted.

As was the case for Experiment 1, response latency analyses were conducted on both raw latencies and z-score transformed latencies. Each type of latency was analyzed using a 2 (print exposure: high, low) × 2 (BOI: high, low) mixed-model ANOVA, with both subject (F_1) and item (F_2) analyses conducted. The mean raw and z-score transformed response latencies for correct responses for the experimental words are presented in **Table 3**. The mean response latencies for correct responses for the pseudohomophones were 874 ms for the high print exposure group and 974 ms for the low print exposure group. The pseudohomophone and non-word error percentages for the high print exposure group were 3.9% and 8.2%, respectively; for the low print exposure group they were 4.4% and 10.2%, respectively.

Raw response latency analysis

For the raw response latency data, there was an effect of print exposure, $F_1(1, 55) = 5.28$, MSE $= 12, 221.37$, $\eta^2 = 0.09$; $F_2(1, 46) = 111.57$, MSE $= 521.38$, $\eta^2 = 0.71$, with the high print exposure group responding on average 47 ms faster than the low print exposure group. There was an effect of BOI, $F_1(1, 55) = 6.09$, MSE $= 838.67$, $\eta^2 = 0.10$; $F_2(1, 46) = 3.43$, $p = 0.07$, MSE $= 1326.45$, $\eta^2 = 0.07$, with responses to high BOI words on average 13 ms faster than responses to low BOI words. There was no interaction between print exposure and BOI, $F_1(1, 55) = 1.76$, $p = 0.19$, MSE $= 838.67$; $F_2(1, 46) = 2.27$, $p = 0.14$, MSE $= 521.38$. Planned comparisons were conducted to examine the effects of BOI for each print exposure group separately. For the high print exposure group, the 21 ms BOI effect was significant, $t_1(28) = 2.69$, SEM $= 7.64$, $\eta^2 = 0.21$; $t_2(46) = 2.58$, SEM $= 8.07$, $\eta^2 = 0.13$. For the low print exposure group, the 6 ms BOI effect was not significant, both ts < 1.

Table 3 | Mean raw response latencies (in ms) and standard errors, mean error percentages and standard errors, and mean transformed response latencies (in *Z*-scores) and standard errors for Experiment 2.

Word type	High print exposure		Low print exposure	
	M	SE	*M*	SE
RAW RESPONSE LATENCIES				
High BOI	605	14.5	660	14.5
Low BOI	626	16.3	666	15.1
BOI effect	+21		+6	
TRANSFORMED RESPONSE LATENCIES				
High BOI	−0.083	0.03	−0.014	0.03
Low BOI	0.082	0.03	0.013	0.03
BOI effect	0.165		0.027	

Note: BOI, body-object interaction.

Z-score transformed response latency analysis

For the z-score transformed response latency data, there was, of course, no effect of print exposure, both Fs < 1. There was an effect of BOI, $F_1(1, 55) = 4.63$, MSE $= 0.06$, $\eta^2 = 0.08$; $F_2(1, 46) = 3.58$, $p = 0.07$, MSE $= 0.06$, $\eta^2 = 0.07$, with a z-score difference between the high BOI words and the low BOI words of 0.096 (and the high BOI words having faster latencies). There was no interaction between print exposure and BOI, $F_1(1, 55) = 2.42$, $p = 0.13$, MSE $= 0.06$; $F_2(1, 46) = 2.72$, $p = 0.11$, MSE $= 0.03$. Planned comparisons were conducted to examine the effects of BOI for each print exposure group separately. For the high print exposure group, the z-score difference between the high BOI words and the low BOI words was 0.165 and was significant, $t_1(28) = 2.47$, SEM $= 0.07$, $\eta^2 = 0.18$; $t_2(46) = 2.58$, SEM $= 0.06$, $\eta^2 = 0.13$. For the low print exposure group, the z-score difference between the high BOI words and the low BOI words was 0.027 and was not significant, both ts < 1.

Cross task analysis

In order to statistically evaluate the opposite patterns observed across our two tasks, we also analyzed the data from both tasks together, in a 2 (print exposure: high, low) × 2 (BOI: high, low) × 2 (task: SCT, PLDT) mixed-model ANOVA. Both subject (F_1) and item (F_2) analyses were conducted. In the subject analyses, print exposure and task were between-subjects variables and BOI was a within-subjects variable. In the item analyses, print exposure and task were within-items variables and BOI was a between-items variable. We report only whether the three-way interaction between print exposure, BOI, and task was significant.

For the raw response latency analysis, the three-way interaction was significant, $F_1(1, 99) = 6.24$, MSE $= 973.39$, $\eta^2 = 0.06$; $F_2(1, 91) = 7.26$, MSE $= 1051.41$, $\eta^2 = 0.07$. For the z-score transformed response latency analysis, the three-way interaction approached significance, $F_1(1, 99) = 3.28$, $p = 0.07$, MSE $= 0.04$, $\eta^2 = 0.03$; $F_2(1, 91) = 3.95$, $p = 0.05$, MSE $= 0.02$, $\eta^2 = 0.04$. These analyses confirm that there were significantly different interactions of print exposure and BOI in our two experiments: whereas high print exposure readers were faster than low print exposure readers in both tasks, they showed smaller BOI effects (than low print exposure readers) in the semantic categorization task and larger BOI effects (than low print exposure readers) in the phonological lexical decision task.

DISCUSSION

Recall that two outcomes were proposed regarding the possible influence of print exposure on BOI in the phonological lexical decision task. The first proposed outcome was that a larger BOI effect should be observed with the low print exposure readers, because their lower integrity phonological representations should be further away from recognition threshold and would, therefore, benefit more from the greater motor simulation elicited by the high BOI words. The second proposed outcome was that a larger effect of BOI may be observed with the high print exposure readers, because they have more efficient orthographic-to-semantics-to-phonology mappings that would allow them to benefit more from the greater motor simulation elicited by the high BOI words.

The two key results from Experiment 2—that the high print exposure readers responded more quickly and were the only print exposure group to have a facilitatory BOI effect—clearly do not support the first proposed outcome, but instead provide support for the second proposed outcome. That is, these two findings are consistent with the idea that print exposure develops more efficient feedback from semantics to phonology, with the end result of faster settling of phonological representations associated with high BOI words than of phonological representations associated with low BOI words, thus producing faster phonological lexical decision latencies for the former type of words. Furthermore, the findings from the present experiment extend previous work examining the effects of print exposure on orthographic and phonological processing, by suggesting that print exposure also develops more efficient orthographic-to-semantics-to-phonological mappings within the visual word recognition system.

GENERAL DISCUSSION

In the present study we examined the influence of print exposure on BOI in visual word recognition, using the semantic categorization and phonological lexical decision tasks. Our findings suggest that print exposure does influence the BOI effect, and that the manner by which it does so depends on task demands.

Recall that according to the semantic feedback framework, responses in the semantic categorization task are primarily based on semantic processing, and that a facilitatory effect of BOI arises under these conditions because high BOI words elicit richer semantic activation (i.e., richer motor simulations) within the semantic units, which leads to faster settling of semantic representations and faster semantic categorization latencies for these types of words. The results from our raw latency analyses indicated that although there was a facilitatory BOI effect for both print exposure groups, the effect was significantly smaller for the high print exposure group. An extension of the lexical integrity hypothesis (Yap et al., 2009) provides an explanation for why this finding was observed. More specifically, for high print exposure readers, greater reading experience leads to the development of higher integrity semantic representations that are closer to recognition threshold (or, in other words, settle more quickly), and to greater efficiency of semantic processing more generally. This results in less benefit for these readers when motor simulations are used to make semantic categorizations, because less stimulus-driven knowledge is needed to correctly categorize the stimuli, which leads to less of an advantage in the recognition of and responding to high BOI words. Conversely, for low print exposure readers, less reading experience leads to the development of lower integrity semantic representations that are further away from recognition threshold (or, in other words, settle less quickly), and to less efficient semantic processing more generally. This results in more benefit for these readers when motor simulations are used to make semantic categorizations, because more stimulus-driven knowledge is needed to correctly categorize the stimuli, which leads to more of an advantage in the recognition of and responding to high BOI words. In summary, under experimental conditions in which responses are primarily based directly on semantic processing, higher integrity representations and more efficient processing leads to an attenuation of the facilitatory BOI effect.

We noted in the Discussion section of Experiment 1 that there is an alternative viewpoint regarding the interpretation of the print exposure by BOI interaction observed in our semantic categorization results. According to this viewpoint (Faust et al., 1999; Yap et al., 2012a), it is difficult to determine, using raw response latencies, whether the larger facilitatory BOI effect associated with the low print exposure group was due to a greater impact of motor simulation elicited by the high BOI words than the low BOI words for this group of readers (i.e., it is due to processing amount), or if the BOI effect was simply correlated with individual differences in overall response latencies (i.e., it is due to processing speed or, in other words, the low print exposure readers had a larger facilitatory BOI effect simply because they took longer to respond). In order to disentangle these two possibilities, Faust et al. suggested transforming raw response latencies to a common scale, in essence statistically partialling out any effect attributable to processing rate, and only examining any effect that may be attributable to processing amount. If the print exposure by BOI interaction remained after such a transformation of the response latency data, then, according to this viewpoint, the interpretation of differential benefits of motor simulation between the two print exposure groups is valid; otherwise, no such interpretation is warranted. Recall that after z-score transforming our raw response latency data, there was no longer a print exposure by BOI interaction. In other words, the facilitatory BOI effect could be interpreted as being of similar magnitude for both print exposure groups.

The analyses based on the raw semantic categorization latencies and the z-score transformation of those latencies lead to different conclusions. The raw response latency findings support the idea that motor simulation is of differential benefit for low print exposure readers and high print exposure readers, whereas the z-score transformed response latency findings support the idea that there is little, if any, differential benefit of motor simulation between the two print exposure groups. We propose that although there may be instances in which the alternative viewpoint of group by variable interactions as indicative of "spurious overadditive interactions" (Faust et al., 1999, p. 777) is valid there may be instances in which this is an overly restrictive way of interpreting these types of interactions. For example, if there is reason to suspect that the groups under examination may have *qualitative* differences in how they process information (such as comparing older participants vs. younger participants, or brain-injured vs. non-brain-injured participants, on some cognitive task), it is likely appropriate to consider processing rate as a variable having two relatively distinct *kinds* of processing. In these types of cases, partialling out the effects of processing rate would be warranted before interpreting a group by variable interaction. However, if there is reason to suspect that the groups under examination may simply have *quantitative* differences in how they process information (such as comparing two groups of university-based, skilled readers), it is likely inappropriate to consider processing rate as a variable having two relatively distinct *kinds* of processing. In these types of cases, which we suggest is indicative of the present study, partialling out the effects of processing rate would not be necessary to interpret a group by variable interaction. Thus, we conclude that print exposure did differentially influence the effects of BOI for the high print exposure readers and low print exposure readers in Experiment 1, although we acknowledge that some researchers may disagree with this interpretation.

Recall that according to the semantic feedback framework, responses in the phonological lexical decision task are primarily based on phonological processing, and that a facilitatory BOI effect arises in this task because high BOI words elicit greater semantic activation (i.e., richer motor simulations) within the semantic units, which then sends stronger semantic feedback to the phonological units, leading to faster settling on phonological representations and hence faster phonological lexical decision latencies. The results from Experiment 2 indicated that a facilitatory BOI effect was observed only for the high print exposure readers. These results are consistent with the idea that reading experience leads to the development of more efficient orthographic-to-phonological-to-semantic mappings, a novel finding, and when this occurs, the effects of semantic feedback are more beneficial for the high print exposure readers, readers whose visual word recognition processes benefit from the increased semantic feedback to phonology that is elicited by high BOI words. Importantly, these conclusions are immune from the criticism of the alternative viewpoint proposed by Faust et al. (1999), because the high print exposure readers exhibited both faster overall phonological lexical decision latencies, and were the only print exposure group for which a facilitatory BOI effect was observed.

In summary, under experimental conditions in which responses are not based directly on semantic processing, but rather may be modulated by semantic feedback, higher integrity representations (i.e., in the case of our Experiment 2, phonological representations) and more efficient processing (i.e., orthographic-to-semantics-to-phonological processing) leads to an increase of the facilitatory BOI effect.

An important limitation of the present study is the reading measure used. Although ART scores are reliably associated with reading experience (Cunningham and Stanovich, 1990) and vocabulary knowledge (West et al., 1993; Lee et al., 1997), they do not capture either dimension perfectly, and there are more precise ways of measuring lexical expertise. For example, Andrews and colleagues have derived and used individual difference measures in written language proficiency with tests of reading, spelling, and vocabulary, and examined the effects of those differences on a number of aspects of lexical retrieval (e.g., masked neighbor priming, Andrews and Hersch, 2010; masked form priming, Andrews and Lo, 2012). This multidimensional approach to assessing lexical expertise should be adopted in future studies of individual differences and semantic processing.

In conclusion, the results of the present study are consistent with the view that lexical conceptual processing is flexible and dynamic in nature (Kiefer and Pulvermüller, 2012). In order to optimize performance in a task, participants can modulate the processing required. More specifically, as a function of print exposure, there is variance in the degree to which embodied semantic information influences lexical conceptual processing. The effects of the semantic richness variable BOI are smaller for high print exposure readers compared to low print exposure readers under conditions in which responses are based on semantic processing (and the former type of readers have higher integrity semantic representations and more efficient semantic processing), but are larger under conditions in which responses are based on phonological processing but may be modulated by semantic feedback (they have higher integrity phonological representations and more efficient orthographic-to-semantics-to-phonological processing).

ACKNOWLEDGMENTS

This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grants to Paul D. Siakaluk and Penny M. Pexman.

REFERENCES

Andrews, S., and Bond, R. (2009). Lexical expertise and reading skill: bottom-up and top-down processing of lexical ambiguity. *Read. Writ.* 22, 687–711.

Andrews, S., and Hersch, J. (2010). Lexical precision in skilled readers: individual differences in masked neighbor priming. *J. Exp. Psychol. Gen.* 139, 299–318.

Andrews, S., and Lo, S. (2012). Not all skilled readers have cracked the code: individual differences in masked form priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 152–163.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Bennett, S. D. R., Burnett, A. N., Siakaluk, P. D., and Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599 multisyllabic nouns. *Behav. Res. Methods* 43, 1100–1109.

Chateau, D., and Jared, D. (2000). Exposure to print and word recognition processes. *Mem. Cogn.* 28, 143–153.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again.* Cambridge, MA: MIT Press.

Cunningham, A. E., and Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: a quick measure of reading experience. *J. Educ. Psychol.* 82, 733–740.

Cunningham, A. E., and Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Dev. Psychol.* 33, 934–945.

Faust, M. E., Balota, D. A., Spieler, D. H., and Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: implications for group differences in response latency. *Psychol. Bull.* 125, 777–799.

Grèzes, J., and Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12, 1–19.

Hargreaves, I. S., Leonard, G. A., Pexman, P. M., Pittman, D. J., Siakaluk, P. D., and Goodyear, B. G. (2012a). The neural correlates of the body-object interaction effect in semantic processing. *Front. Hum. Neurosci.* 6:22. doi: 10.3389/fnhum.2012.00022

Hargreaves, I. S., Pexman, P. M., Zdrazilova, L., and Sargious, P. (2012b). How a hobby can shape cognition: visual word recognition in competitive Scrabble players. *Mem. Cogn.* 40, 1–7.

Hino, Y., and Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: an alternative to lexical access accounts. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1331–1356.

Hino, Y., Lupker, S. J., and Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: interactions between orthography, phonology, and semantics. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 686–713.

ITP Nelson Canadian Dictionary. (1997). *ITP Nelson Canadian Dictionary of the English Language: An Encyclopedic Reference.* Scarborough, ON: ITP Nelson.

Kiefer, M. and Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex*. [Epub ahead of print].

Lakoff, G., and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought.* New York, NY: Basic Books.

Lee, S-Y., Krashen, S., and Tse, L. (1997). The author recognition test and vocabulary knowledge: a replication. *Percept. Mot. Skills* 85, 1428–1430.

Pecher, D. (2001). Perception is a two-way junction: feedback semantics in word recognition. *Psychon. Bull. Rev.* 8, 545–551.

Pecher, D., and Zwaan, R. A. (2005). "Introduction to grounding cognition," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*, eds D. Pecher and R. A. Zwaan, (Cambridge, UK: Cambridge University Press), 1–7.

Péran, P., Démonet, J. F., Cherubini, A., Carbebat, D., Caltagirone, C., and Sabatini, U. (2010). Mental representations of action: the neural correlates of the verbal and motor components. *Brain Res.* 1328, 89–103.

Perfetti, C. A. (1992). "The representation problem in reading acquisition," in *Reading Acquisition*, eds P. B. Gough, L. C. Ehri, and R. Treiman (Hillsdale, NJ: Erlbaum), 145–174.

Perfetti, C. A., and Hart, L. (2002). "The lexical quality hypothesis," in *Precursors of Functional Literacy*, eds L. Vehoeven, C. Elbro, and P. Reitsma (Amsterdam: John Benjamins), 180–213.

Pexman, P. M., and Lupker, S. J. (1999). Ambiguity and visual word recognition: can feedback explain both homophone and polysemy effects? *Can. J. Exp. Psychol.* 53, 323–334.

Sears, C. R., Siakaluk, P. D., Chow, V. C., and Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect? *J. Psycholinguist. Res.* 37, 269–291.

Siakaluk, P. D., Buchanan, L., and Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Mem. Cogn.* 31, 100–113.

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., and Sears, C. R. (2008a). Evidence for the activation of sensorimotor information during visual word recognition: the body-object interaction effect. *Cognition* 106, 433–443.

Siakaluk, P. D., Pexman, P. M., Dalrymple, H-A. R., Stearns, J., and Owen, W. J. (2011). Some insults are more difficult to ignore: the embodied insult Stroop effect. *Lang. Cogn. Process.* 26, 1266–1294.

Siakaluk, P. D., Pexman, P. M., Sears, C. R., and Owen, W. J. (2007). Multiple meanings are not necessarily a disadvantage in semantic processing: evidence from homophone effects in semantic categorisation. *Lang. Cogn. Process.* 22, 453–467.

Siakaluk, P. D., Pexman, P. M., Sears, C. R., Wilson, K., Locheed, K., and Owen, W. J. (2008b). The benefits of sensorimotor knowledge: body-object interaction facilitates semantic processing. *Cogn. Sci.* 32, 591–605.

Siakaluk, P. D., Sears, C. R., and Lupker, S. J. (2002). Orthographic neighborhood effects in lexical decision: the effects of nonword orthographic neighborhood size. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 661–681.

Stanovich, K. E., and West, R. F. (1989). Exposure to print and orthographic processing. *Read. Res. Q.* 24, 402–433.

Tillotson, S. M., Siakaluk, P. D., and Pexman, P. M. (2008). Body-object interaction ratings for 1,618 monosyllabic nouns. *Behav. Res. Methods* 40, 1075–1078.

Tousignant, C., and Pexman, P. M. (2012). Flexible recruitment of semantic richness: context modulates body-object interaction effects in lexical-semantic processing. *Front. Hum. Neurosci.* 6:53. doi: 10.3389/fnhum.2012.00053

Unsworth, S. J., and Pexman, P. M. (2003). The impact of reader skill on phonological processing in visual word recognition. *Q. J. Exp. Psychol.* 56, 63–81.

Wellsby, M., Siakaluk, P. D., Owen, W. J., and Pexman, P. M. (2011). Embodied semantic processing: the body-object interaction effect in a non-manual task. *Lang. Cogn.* 3, 1–14.

West, R. F., Stanovich, K. E., and Mitchell, H. R. (1993). Reading in the real world and its correlates. *Read. Res. Q.* 28, 34–50.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.

Yap, M. J., Balota, D. A., Sibley, D. E., and Ratcliff, R. (2012a). Individual differences in visual word recognition: insights from the English Lexicon Project. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 53–79.

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., and Huff, M. J. (2012b). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:72. doi: 10.3389/fnhum.2012.00072

Yap, M. J., Tse, C-S., and Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency revealed by RT distributional analyses: the role of lexical integrity. *J. Mem. Lang.* 61, 303–325.

Zwaan, R. A. (2008). "Experiential traces and mental simulations in language comprehension," in *Symbols and Embodiment: Debates on Meaning and Cognition*, eds M. de Vega, A. M. Glenberg, and A. C. Graesser, (New York, NY: Oxford University Press), 165–180.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 January 2012; accepted: 13 April 2012; published online: 03 May 2012.

Citation: Hansen D, Siakaluk PD and Pexman PM (2012) The influence of print exposure on the body-object interaction effect in visual word recognition. Front. Hum. Neurosci. ***6****:113. doi: 10.3389/fnhum.2012.00113*

APPENDIX

ITEMS USED IN THE EXPERIMENTS

High BOI words

belt, brick, couch, crown, crumb, dish, drum, fence, flute, gift, grape, lamp, mask, pear, pipe, purse, rope, skirt, stool, suit, tape, thorn, tool, vest

Low BOI words

cake, cliff, cloud, clown, creek, dirt, ditch, dorm, flame, flood, juice, kite, lace, leaf, mist, pond, seed, shelf, ship, silk, smog, torch, tribe, tube

Less imageable words

chasm, clout, cusp, farce, fare, fate, fault, feat, flaw, fleck, fluke, fraud, froth, gist, hint, hoax, lack, lapse, loss, luck, noun, oath, pact, pang, phase, plea, ploy, pride, proof, prose, realm, risk, sake, scorn, sect, skill, soul, span, spoof, tact, trait, trend, truce, trust, verb, whiff, whim, zeal

Pseudohomophones

berd, boal, boan, bote, brane, crain, dait, doar, drane, gaim, gard, goast, gote, groop, gurl, hoam, hoap, hoze, jale, jerm, jirk, joak, klaim, koast, nale, noat, nurve, rane, rong, rore, roze, scail, sheat, shurt, skalp, skarf, sleap, smoak, stawl, stoar, swet, teath, thret, tode, treet, tutch, werk, wheet

Non-words

bame, beal, besh, bime, binch, bope, bram, brame, brank, brate, bulch, chate, cheen, clace, clirp, crong, cruss, dack, dake, dawk, dreeb, dunch, duss, fage, filt, fitch, flane, flang, flef, flet, foom, fulk, fung, gake, gick, glank, gless, grabe, grafe, gurse, hain, hape, hean, helt, hife, hine, jick, jote, kine, kooce, loke, ludge, meep, merch, moach, nent, nerbe, pake, pame, pape, pell, petch, pilk, pleap, poote, potch, pribe, prog, pung, rame, rask, rell, scaff, scug, shate, shink, slirt, soat, spale, spen, spoop, stort, strup, tain, talt, tane, tark, thurn, tinch, toin, trake, treen, trine, turt, vank, yelf

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 15 March 2012
doi: 10.3389/fnhum.2012.00053

Flexible recruitment of semantic richness: context modulates body-object interaction effects in lexical-semantic processing

*Cody Tousignant and Penny M. Pexman**

Department of Psychology, University of Calgary, Calgary, AB, Canada

Edited by:
Paul D. Siakaluk, University of Northern British Columbia, Canada

Reviewed by:
Milena Rabovsky, Humboldt-Universität zu Berlin, Germany
Ron Borowsky, University of Saskatchewan, Canada

****Correspondence:***
Penny M. Pexman, Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada.
e-mail: pexman@ucalgary.ca

Body-object interaction (BOI) is a semantic richness variable that measures the perceived ease with which the human body can physically interact with a word's referent. Lexical and semantic processing is facilitated when words are associated with relatively more bodily experience. To date, BOI effects have only been examined in the context of one semantic categorization task (SCT; is it imageable?). It has been argued that semantic processing is dynamic and can be modulated by context. We examined these influences by testing how task knowledge modulated BOI effects. Participants discriminated between the same sets of entity (high- and low-BOI) and action words in each of four SCTs. Task framing was manipulated: participants were told about one (is it an action? vs. is it an entity?) or both (action or entity? vs. entity or action?) categories of words in the decision task. Facilitatory BOI effects were only observed when participants knew that "entity" was part of the decision category. That BOI information was only useful when participants had expectations that entity words would be presented suggests a strong role for the decision context in lexical-semantic processing, and supports a dynamic view of conceptual knowledge.

Keywords: semantic richness, semantic categorization task, body-object interaction, lexical-semantic, task effects

The study of semantic richness effects has provided valuable insight into the process by which meaning is derived from words. The fact that lexical-semantic processing is facilitated when words have relatively more semantic neighbors (Buchanan et al., 2001), or relatively more features (Pexman et al., 2002, 2003; Grondin et al., 2009), or evoke more imagery (Balota et al., 2004), or more bodily experience (Siakaluk et al., 2008a,b; Tillotson et al., 2008) suggests that these dimensions are all relevant to semantic processing and, presumably, to semantic representation.

For instance, recent research by Siakaluk and colleagues has examined how sensorimotor experience is relevant to lexical-semantic processing. More specifically, as a counterpoint to semantic richness variables such as imageability which focus on sensory experience, Siakaluk and colleagues were interested in the extent to which subjects' motor interactions with a word's referent affected lexical-semantic processing. To this end, Siakaluk et al. (2008a) collected body-object interaction (BOI) ratings for a series of words by having participants rate how easily they could interact with each word's referent. They then presented participants with low-BOI words (i.e., referents were relatively hard to interact with, e.g., *ship*) or high-BOI words (i.e., referents were relatively easy to interact with, e.g., *belt*) in a lexical decision task (LDT) as well as SCT with the "is it imageable?" decision category (Siakaluk et al., 2008b; Wellsby et al., 2011). In both tasks, high-BOI words were classified faster than low-BOI words. These BOI effects have been interpreted in the same framework as many other richness effects (e.g., imageability, number of features, semantic neighborhood density). That is, richer concepts generate stronger semantic activation, facilitating SCT performance, and also providing stronger feedback to orthographic units, facilitating LDT performance.

A lingering theoretical question about BOI and other richness effects, however, is whether these dimensions are a function of stable or dynamic semantic representations. Traditionally, the representation of conceptual knowledge has been characterized as stable and invariant (e.g., Collins and Loftus, 1975; Fodor, 1975). That is, the process of generating word meaning involves activation of a fixed set of properties or features and this process is not modified by task demands. Alternatively, it has been argued that the process of generating word meaning is at least to some degree context dependent (e.g., Barsalou, 1982), such that context determines the particular features activated (Hoenig et al., 2008; Kiefer and Pulvermüller, 2012). This distinction is present in more recent theories, with some proposing stable conceptual representations (e.g., Caramazza and Mahon, 2003) and others proposing more flexibility (e.g., Barsalou, 2008). The goal of the present study was to investigate whether task context modulates the effect of semantic richness, in order to establish how flexibly this information is used.

There is some evidence from previous studies that semantic richness effects vary across tasks. When several different semantic richness effects were compared across LDT and SCT, Pexman et al. (2008) reported that while the effects of contextual dispersion and number of features were significant in both tasks, the effect of number of semantic neighbors was significant only in LDT. Similarly, Yap et al. (2011) compared semantic richness effects across naming, LDT, and SCT, and reported that, while some richness effects were observed across all tasks (e.g., contextual dispersion, number of features), semantic neighborhood density was significant only in LDT. Based on their findings, these authors have argued that readers can dynamically adjust the kinds of information they use to suit the specific demands of the tasks that they face; depending upon a reader's goals they might emphasize some dimensions at the expense of others.

Other studies have compared richness effects within the SCT, by contrasting the effects observed under different decision categories. Pexman et al. (2003) examined how the breadth of a decision category affected categorization latencies for a critical set of items in SCT. Participants categorized low- and high-number of features birds, as well as fillers, in one of three decision categories. The categories ranged from specific to broad: is it a bird? Is it a living thing? Is it concrete? Number of features effects were observed in all three tasks, reflecting faster categorization latencies for the high- versus low-number of features birds, but the size of the number of features effect was approximately twice as large when the decision category was broad (concrete vs. abstract) than when it was narrow (bird vs. nonbird). In a similar vein, semantic ambiguity effects were observed for broad, but not narrow, decision contexts in a SCT (Hino et al., 2006). Hino and colleagues argued that with a narrow decision category participant make their decision by checking a small number of candidate features for each word, whereas with a broad decision category participants invoke more analytic processing to evaluate all activated features for each word. For example, when deciding whether a word refers to a bird, participants can focus on diagnostic semantic information such as whether the referent has feathers or a beak, while ignoring irrelevant semantic information such as how imageable the referent may be. This is not possible for more general decision contexts, where a wide array of semantic features is relevant to the category judgment.

Relatedly, Hargreaves et al. (2012) recently used fMRI to compare the neural correlates of two SCT conditions. The two SCTs involved different decision categories: *is it an animal?* vs. *is it a concrete thing?* Participants completed both tasks and, across participants, the same core set of items were presented in both tasks. The fMRI results showed relatively more activity for the animal SCT in cortical regions that have been linked to general knowledge (e.g., left superior and middle temporal gyri) while the concrete SCT showed relatively more activity in motor regions. These results are interpreted as evidence of top-down modulation of semantic processing; participants make adjustments to optimize performance in a given task and these adjustments have consequences for the activation observed.

The present experiment provides a conceptual extension of recent work examining the role of the decision context in modulating effects of semantic richness (Pexman et al., 2003; Hino et al., 2006), but here we used a more fine-grained manipulation of task context than has been achieved in previous studies. In the previous studies examining decision context effects in SCT, different items (fillers) have been included in each condition. That is, while the analyses have focused on a core set of items presented in every condition, the other items in each condition differed. Thus, the decision context was not the only thing that varied across conditions. In the present study, however, we used exactly the same item sets in every condition. The only thing that varied was the way the decision was presented to participants. Here the context manipulation was not of category breadth but rather category framing. In each of four different decision contexts, participants discriminated between the same sets of object words (concerned that the term "object" would be interpreted too narrowly we referred to these as "entity" words, and they included both low- and high-BOI words), and action words. The BOI dimension is particularly well-suited to this framing manipulation because it is relevant to a certain class of words: concrete nouns. We expected that in general high-BOI words would be classified more quickly and accurately than low-BOI words. The decision context was manipulated by varying the information participants were given about the types of items in the decision task (see **Table 1**).

We manipulated whether participants were told about one (i.e., is it an entity? vs. is it an action?) or both (entity or action? vs. action or entity?) categories of words present in the task. This allowed us to assess whether BOI effects depend on participants knowing that object information will be relevant to the task. If BOI effects are ubiquitous to semantic processing, then this manipulation should have no effect, and a BOI effect should be observed in every version of the decision category. If, on the other hand, semantic processing is modulated by participants' expectations about the relevant information in a task, then BOI effects should not be observed in every version of the decision category and, in particular, may be attenuated in the "is it an action?" version of the decision, where participants are not told in advance that object words will be presented for categorization.

METHOD

PARTICIPANTS

One hundred and fifty-nine University of Calgary undergraduate psychology students were randomly assigned to the entity ($n = 41$), entity-action ($n = 39$), action-entity ($n = 39$), or action ($n = 40$) decision conditions and participated in exchange for course credit. All participants had normal or corrected to normal vision and were fluent English speakers.

Table 1 | Study design.

Instruction condition	*n*	Participants' expectations	Left response button	Right response button
Is it an entity?	41	Entity words and non-entity words	Low- and high-BOI words	Action words
Is it an entity or an action?	39	Entity words and action words	Low- and high-BOI words	Action words
Is it an action or an entity?	39	Action words and entity words	Action words	Low- and high-BOI words
Is it an action?	40	Action words and non-action words	Action words	Low- and high-BOI words

STIMULI

The BOI ratings collected by Tillotson et al. (2008) were used to compile lists of potential low- and high-BOI entity words. In addition, a list of potential action words (e.g., *jump*) was selected from the MRC database (Wilson, 1988).

To ensure that the entity words were conceptually distinct from the action words, a separate group of 45 participants used a six-point Likert scale to rate how action-like the potential stimuli were (1 = entity, 6 = action). This information was used to compile a final list of 35 low- and 35 high-BOI words which had all received low ratings on the entity-action scale, and were matched on a number of dimensions (see **Table 2**). Finally, 70 action words that received high ratings on the entity-action scale were selected. All items are listed in the Appendix.

PROCEDURE

Stimuli were presented using E-Prime software (Schneider et al., 2002) on a 20″ CRT monitor. On each trial, a 500 ms fixation cross was presented, followed by a 60 ms blank screen. This was followed by a target word. In the entity condition, participants were asked to press the left button on a response pad for words that referred to entities and to press the right button for words that referred to non-entities. In the entity-action condition, participants were asked to press the left button in response to words that referred to entities and use the right button to respond to words that referred to actions, whereas in the action-entity condition, participants responded with the left button for words that referred to actions and the right button for words that referred to entities. Finally, in the action condition, participants were asked to provide a left button response to words that referred to actions and a right button response to words that referred to non-actions. The decision category (i.e., "Is it an Entity?" "Entity or Action?" "Action or Entity?" or "Is it an Action?") was presented above each target word. Participants were asked to respond as quickly and as accurately as possible.

To ensure that participants understood the instructions, the experimenter remained in the room while six practice trials were presented. Once the practice trials were completed, the experimenter left the room and the participants completed the remaining 140 experimental trials.

RESULTS

The data for one participant from the action condition and three participants from the entity condition were removed from the final analysis due to poor categorization performance for the critical words (accuracy < 70%). The final group sizes were thus 38 in the entity condition and 39 in the remaining three conditions.

The data for any item for which participants demonstrated less than 70% categorization accuracy were removed from the analysis: *fog*, *back*, *well*, *song*, and *case* (all low-BOI items) in the entity condition (3.57% of the data), *flea* (low BOI), *pat*, and *sue* (both actions) in the entity-action condition (2.15% of the data), *flea* (low BOI) in the action-entity condition (0.07% of the data), and *boot* (high BOI) and *flea* (low BOI) in the action condition (1.43% of the data). In addition, trials with response latencies faster than 350 ms or slower than 2500 ms were removed from the analysis (entity condition: 0.34% of the data; entity-action condition: 3.15% of the data; action-entity condition: 3.90% of the data; action condition: 1.55% of the data).

Mean RTs and accuracy (see **Table 3**) were analyzed with four (decision category: entity, entity-action, action-entity, action) × 2 (BOI: low, high) mixed factors ANOVAs by subjects (F_1) and by items (F_2). In the subject analysis, condition was the between-subjects variable and BOI was the within-subjects variable. In the item analysis, condition was the within-item variable and BOI was the between-item variable.

RT ANALYSIS

In the analysis of RT data, there was an interaction of decision category and BOI, $F_1(3, 151) = 10.27$, $p < 0.001$, $\eta^2 =$

Table 2 | Mean (SD) characteristics of low-BOI and high-BOI word stimuli.

Characteristic	Low BOI	High BOI	*p*
BOI rating	3.39 (0.55)	5.67 (0.46)	<0.001
Entity-action rating	1.69 (0.30)	1.60 (0.25)	0.20
Word length	4.14 (0.84)	4.14 (0.84)	1.00
Familiarity	5.32 (0.43)	5.32 (1.03)	0.99
Imageability	5.72 (0.39)	5.69 (0.30)	0.75
Concreteness	5.67 (0.38)	5.76 (0.34)	0.33
Orthographic neighbors	9.08 (5.54)	8.94 (6.00)	0.91
Kucera-Francis frequency	96.31 (219.79)	87.08 (212.15)	0.93
CELEX frequency	116.08 (303.82)	95.41 (202.19)	0.73
Standard frequency index	52.74 (8.92)	53.59 (6.02)	0.64
Bigram frequency	1841.62 (926.46)	1678.30 (859.20)	0.36
Contextual dispersion	0.68 (0.15)	0.71 (0.14)	0.43

Note: BOI ratings were taken from Tillotson et al. (2008) norms. Entity Ratings ranged from 1 (entity) – 6 (action) and were collected in a pilot study. Familiarity, imageability and concreteness measures were taken from the MRC Psycholinguistic Database [Wilson (1988)], Orthographic neighbors and bigram frequency measures were taken from the English Lexicon Project [Balota et al. (2007)]. Kucera-Francis Frequency [Kucera and Francis (1967)]. CELEX Frequency = Dutch Centre for Lexical Information frequency measure [Davis (2005)]. Standard Frequency Index and Contextual Dispersion measures taken from the Educator's Word Frequency Guide [Zeno et al. (1995)].

Table 3 | Mean (SD) RT and accuracy for low-BOI words, high-BOI words, and action words.

Decision category	Low BOI words		High BOI words		BOI effect		Action words	
	RT	Accuracy	RT	Accuracy	RT	Accuracy	RT	Accuracy
Entity	1031 (168)	0.87 (0.11)	906 (146)	0.95 (0.03)	−125	0.08	1021 (178)	0.90 (0.10)
Entity-action	968 (202)	0.93 (0.09)	911 (169)	0.93 (0.07)	−57	0.005	874 (136)	0.92 (0.06)
Action-entity	1046 (182)	0.95 (0.05)	995 (171)	0.95 (0.04)	−51	−0.005	971 (145)	0.92 (0.05)
Action	859 (147)	0.96 (0.04)	851 (146)	0.95 (0.04)	−8	−0.01	812 (125)	0.91 (0.08)

0.12; $F_2(3, 183) = 9.31$, $p < 0.001$, $\eta^2 = 0.07$. High-BOI words were categorized faster than low-BOI words in the entity, $t_1(37) = 9.18$, $p < 0.001$, $\eta^2 = 0.69$; $t_2(50.11) = 4.34$, $p < 0.001$, $\eta^2 = 0.27$, entity-action, $t_1(38) = 3.60$, $p = 0.001$, $\eta^2 = 0.25$; $t_2(67) = 2.14$, $p < 0.05$, $\eta^2 = 0.06$, and action-entity, $t_1(38) = 2.70$, $p = 0.01$, $\eta^2 = 0.16$; $t_2(67) = 2.28$, $p < 0.05$, $\eta^2 = 0.07$, conditions. Critically, there was no significant difference between categorization times for high- and low-BOI items in the action condition, $t_1 < 1$; $t_2 < 1$, suggesting that BOI information was only useful when participants were told that entity words would be present in the decision task[1].

Results also included a significant main effect of BOI, as high-BOI words were classified more quickly than low-BOI words overall, $F_1(1, 151) = 64.44$, $p < 0.001$, $\eta^2 = 0.29$; $F_2(1, 61) = 10.50$, $p < 0.01$, $\eta^2 = 0.14$. There was also a significant main effect of decision category, $F_1(3, 151) = 7.23$, $p < 0.001$, $\eta^2 = 0.13$; $F_2(3, 183) = 63.22$, $p < 0.001$, $\eta^2 = 0.47$, since RTs to the critical words were fastest for the action condition (855 ms), followed by the entity-action condition (939 ms), entity condition (969 ms), and action-entity condition (1021 ms). This effect of decision category was followed up using comparisons with a Bonferroni correction. RTs in the action condition were faster than those in the entity, $t_1(75) = 3.37$, $p < 0.01$; $t_2(62) = 7.14$, $p < 0.01$, and action-entity conditions, $t_1(76) = 4.72$, $p < 0.01$; $t_2(66) = 15.37$, $p < 0.01$. There was no significant difference between RTs in the action and entity-action conditions in the subject analysis, $t_1(76) = 2.29$, $p > 0.05$, however, RTs were faster in the action than entity-action condition in the item analysis $t_2(66) = 7.54$, $p < 0.01$. RTs in the entity-action condition were not found to differ from RTs in the entity condition in the subject analysis, $t_1 < 1$, but entity-action RTs were significantly faster than the entity RTs in the item analysis, $t_2(64) = 3.18$, $p < 0.05$. Entity-action RTs did not differ from action-entity RTs in the subject analysis, $t_1(76) = 2.08$, $p > 0.05$, but entity-action RTs were significantly faster than action-entity RTs in the item analysis $t_2(68) = 9.33$, $p < 0.01$. Finally, RTs in the entity and action-entity conditions did not differ, $t_1 < 1$; $t_2(64) = 2.81$, $p > 0.05$.

[1]Although the difference is not significant, action-entity ratings were numerically lower (i.e., more entity-like) for high- versus low-BOI words. To investigate whether this difference influenced the overall pattern of results, we re-examined the data using ANCOVA analyses in which action-entity rating was a covariate. With one exception, the pattern of results revealed by these new analyses was identical to that of the reported analyses: in the original analysis of RTs in the entity-action condition, a significant BOI advantage was observed (faster RTs for high- vs. low-BOI words). This same effect was marginally significant ($p = 0.062$) in the ANCOVA analysis.

ACCURACY ANALYSIS

In the analysis of accuracy data, decision category and BOI were again found to interact, $F_1(3, 151) = 14.17$, $p < 0.001$, $\eta^2 = 0.21$; $F_2(3, 183) = 19.20$, $p < 0.001$, $\eta^2 = 0.20$. There was no difference in categorization accuracy for high- versus low-BOI items in the entity-action, $t_1 < 1$; $t_2 < 1$, and action-entity conditions, $t_1 < 1$; $t_2 < 1$. In the action task, high-BOI items were categorized marginally less accurately than low-BOI words in the subject analysis, $t_1(38) = 1.81$, $p = 0.07$, $\eta^2 = 0.07$, but not the item analysis, $t_2(65) = 1.01$, $p = 0.31$, $\eta^2 = 0.01$. Finally, high-BOI words were categorized more accurately than low-BOI words in the entity condition, $t_1(37) = 9.18$, $p < 0.001$, $\eta^2 = 0.69$; $t_2(54.04) = 5.32$, $p < 0.001$, $\eta^2 = 0.34$, suggesting that BOI information was most useful when participants were only told that entity words would be present in the decision task.

Results also included a main effect of BOI, as high-BOI words were categorized more accurately than low-BOI words overall, $F_1(1, 151) = 9.10$, $p < 0.01$, $\eta^2 = 0.05$; $F_2(1, 61) = 3.84$, $p = 0.055$, $\eta^2 = 0.05$. There was also a main effect of decision category, $F_1(3, 151) = 4.86$, $p < 0.01$, $\eta^2 = 0.09$; $F_2(3, 183) = 17.37$, $p < 0.001$, $\eta^2 = 0.18$, with the highest accuracy rates for critical words in the action condition (0.96), followed by the action-entity condition (0.95), the entity-action condition (0.93) and the entity condition (0.91). Follow-up comparisons using the Bonferroni correction revealed that accuracy was significantly higher in the action condition than the entity condition, $t_1(60.50) = 3.49$, $p < 0.01$; $t_2(62) = 3.98$, $p < 0.01$, but not the action-entity condition, $t_1 < 1$; $t_2 < 1$. There was no significant difference between accuracy in the action versus entity-action conditions in the subject analysis, $t_1(59.27) = 2.16$, $p > 0.05$, however, responses were significantly more accurate in the action condition than the entity-action condition in the item analysis, $t_2(66) = 4.15$, $p < 0.01$. Responses were significantly more accurate in the action-entity condition than the entity condition, $t_1(61.47) = 2.96$, $p < 0.05$; $t_2(64) = 3.65$, $p < 0.01$. Accuracy was not found to differ between the action-entity and entity-action conditions in the subject analysis, $t_1(60.20) = 1.68$, $p > 0.05$, however, responses in the action-entity condition were found to be significantly more accurate than those in the entity-action condition in the item analysis, $t_2(68) = 3.48$, $p < 0.01$. Finally, accuracy did not differ across the entity-action and entity conditions, $t_1 < 1$; $t_2(64) = 1.45$, $p > 0.05$.

DISCUSSION

The goal of the current study was to examine how task context might modulate BOI effects in lexical-semantic processing. As such, we investigated how small differences in task presentation (participants' advance knowledge about the types of items presented) modulated BOI effects in semantic categorization behavior. Results showed a BOI advantage, with faster classification times for high- versus low-BOI words, in the three conditions where participants were told to expect entity words. This effect was largest, and was accompanied by a significant accuracy advantage for high BOI words, when participants were only expecting entity (and non-entity) words. Critically, there was no BOI advantage for either RT or accuracy measures when participants were only expecting action (and non-action) words. Our results show three quantitatively different effects using the same set of items, indicating strong modulation of BOI effects as a function of the specific task context.

That the effect of BOI is modulated by the information participants are given about the decision category indicates a strong role for context in semantic processing and suggests that participants are able to adopt disparate task sets as a function of their expectations about what will be important in a task. Certain category labels bias participants' response behavior, perhaps by encouraging them to focus on dimensions of semantic information that are highly relevant to the decision. In the entity condition, this focus on entity-relevant dimensions magnified the BOI effect. Conversely, in the action condition, the focus on action-relevant dimensions eliminated the BOI effect. When participants were given information about both categories of words, as in the action-entity and entity-action conditions, a middle ground was reached, where it seems probable that dimensions relevant to both categories were emphasized in order to categorize stimuli. As such, our results are consistent with those of other studies showing that the type of information participants extract from word stimuli depends on the task context. For instance, Raposo et al. (2009) showed that auditory processing of action verbs was associated with motor and premotor activation when the words were presented alone or in literal sentences, but not when the words were intended figuratively, in idiomatic sentences (although cf. Boulenger et al., 2009).

The observation of BOI effects in lexical-semantic processing has been taken as evidence for the claim that bodily experience is an important aspect of semantic knowledge, activated in the process of generating meaning from print (e.g., Siakaluk et al., 2008a,b; Wellsby et al., 2011). The absence of a significant BOI effect in the "is it an action?" condition of the present study puts limits on the impact of this bodily experience dimension in semantic processing. This is not necessarily to say that some semantic processing is disembodied, but rather highlights the fact that a given embodied dimension (e.g., BOI) may only influence behavior in contexts in which it is task-relevant (Willems and Casasanto, 2011). Semantic processing in the "is it an action?" condition may still be grounded in sensorimotor processing, but whatever sensorimotor dimensions are most relevant to the action decision category were not captured in the present study.

We chose to examine BOI effects in the present research because BOI is a richness dimension that is relevant to a particular class of words. As such, it seemed possible that BOI might be sensitive to the kind of task framing manipulation applied here, where participants had advance knowledge that a type of word would be presented (or not). An unanswered question, however, is whether this kind of task context modulation would also be observed for other semantic richness effects. That is, are the effects of other measures of semantic richness on semantic categorization performance equally malleable? Some richness dimensions, like BOI and imageability, are derived from subjective ratings and, arguably, are thus more intuitive than other objectively-derived dimensions, like number of features, semantic neighborhood density, or contextual dispersion. That is, people can provide consistent ratings about words' BOI but our experience with the more objectively-derived number of features dimension, for instance, is that people have very little insight about whether a word has a high or low number of features. One possibility is that context effects are strongest when participants have some insight about the kinds of information associated with particular stimuli, and can use this insight to tap into the information they suspect will optimize their performance in a task. If this is the case, then the subjective semantic richness dimensions may be more malleable than the objective richness dimensions. Of course, it is also possible that this insight is not at all relevant to context modulation, and all semantic dimensions are equally context-dependent. In this latter case it should not matter whether the particular richness dimension is subjectively- or objectively-derived. These possibilities will need to be tested in future research.

The results of the present study show that even quite subtle differences in the way a task is characterized can produce substantial changes in behavioral effects, as participants made adjustments to their processing of word meaning information (based entirely on small changes in the instructions provided) in order to optimize performance in categorization tasks. As such, our results suggest that participants have strong top-down control of the semantic categorization process. This suggestion is compatible with a flexible, dynamic view of semantic processing (Kiefer and Pulvermüller, 2012).

ACKNOWLEDGMENTS

The Natural Sciences and Engineering Research Council of Canada supported this research through a Postgraduate Scholarship to CT and a Discovery Grant to Penny M. Pexman.

REFERENCES

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. J. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. Gen.* 133, 283–316.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behav. Res. Methods* 39, 445–459.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Mem. Cognit.* 10, 82–93.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Boulenger, V., Hauk, O., and Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914.

Buchanan, L., Westbury, C., and Burgess, C. (2001). Characterizing semantic space: neighborhood effects in word recognition. *Psychon. Bull. Rev.* 8, 531–544.

Caramazza, A., and Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends Cogn. Sci.* 7, 354–361.

Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428.

Davis, C. J. (2005). N-Watch: a program for deriving neighborhood size and other pshycolinguistic statistics. *Behav. Res. Methods* 37, 65–70.

Fodor, J. A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Grondin, R., Lupker, S. J., and McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *J. Mem. Lang.* 60, 1–19.

Hargreaves, I. S., White, M., Pexman, P. M., Pittman, D., and Goodyear, B. G. (2012). The question shapes the answer: the neural correlates of task differences reveal dynamic semantic processing. *Brain Lang.* 120, 73–78.

Hino, Y., Pexman, P. M., and Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: are they due to semantic coding? *J. Mem. Lang.* 55, 247–273.

Hoenig, K., Sim, E-J., Bochev, V., Herrnberger, B., and Kiefer, M. (2008). Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *J. Cogn. Neurosci.* 20, 1799–1814.

Kiefer, M., and Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex*. (in press).

Kucera, H., and Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Press.

Pexman, P., Hargreaves, I., Siakaluk, P., Bodner, G., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on visual word recognition. *Psychon. Bull. Rev.* 15, 161–167.

Pexman, P. M., Holyk, G. G., and Monfils, M-H. (2003). Number-of-features effects and semantic processing. *Mem. Cognit.* 31, 842–855.

Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number-of-features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.

Raposo, A., Moss, H. E., Stamatkis, E. A., and Tyler, L. K. (2009). Modulation of motor and premotor cortices by actions, action words, and action sentences. *Neuropsychologia* 47, 388–396.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime Reference Guide*. Sharpsburg, PA: Psychology Software Tools, Inc.

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., and Sears, C. R. (2008a). Evidence for the activation of sensorimotor information during visual word recognition: the body-object interaction effect. *Cognition* 106, 433–443.

Siakaluk, P. D., Pexman, P. M., Sears, C., Wilson, K., Locheed, K., and Owen, W. J. (2008b). The benefits of sensorimotor knowledge: body-object interaction facilitates semantic processing. *Cogn. Sci.* 32, 591–605.

Tillotson, S. M., Siakaluk, P. D., and Pexman, P. M. (2008). Body-object interaction ratings for 1,618 monosyllabic nouns. *Behav. Res. Methods* 40, 1075–1078.

Wellsby, M., Siakaluk, P. D., Owen, W. J., and Pexman, P. M. (2011). Embodied semantic processing: the body-object interaction effect in a non-manual task. *Lang. Cogn.* 3, 1–14.

Willems, R. M., and Casasanto, D. (2011). Flexibility in embodied language understanding. *Front. Psychol.* 2:116. doi: 10.3389/fpsyg.2011.00116

Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behav. Res. Methods Instrum. Comput.* 20, 6–11.

Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.

Zeno, S. M., Ivens, S. H., Millard, R. T., and Duvvuri, R. (1995). *The Educator's Word Frequency Guide*. United States of America: Touchstone Applied Science Associates, Inc.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 December 2011; accepted: 29 February 2012; published online: 15 March 2012.

Citation: Tousignant C and Pexman PM (2012) Flexible recruitment of semantic richness: context modulates body-object interaction effects in lexical-semantic processing. Front. Hum. Neurosci. ***6****:53. doi: 10.3389/fnhum.2012.00053*

APPENDIX

ITEMS USED IN THE EXPERIMENT

Low-BOI words

ash, back, band, bay, birch, brain, brass, case, flea, fog, frost, gang, hall, heart, jail, king, knight, lane, lint, lung, mink, pint, pit, prince, roof, song, spade, star, stripe, sun, tail, tar, tomb, well, zoo

High-BOI words

belt, boot, bowl, cage, cart, child, couch, feet, food, friend, gate, gift, ham, hat, ice, man, mat, mole, nail, neck, palm, pearl, pie, priest, purse, room, seat, silk, stair, string, suit, toy, tube, vest, wheel

Action words

act, assist, attend, beg, blow, build, bury, carry, choose, come, cope, draw, earn, eat, escape, fight, gasp, give, glare, hear, hide, hold, hunt, ignore, jump, kick, kill, kiss, lean, learn, lift, make, manage, meet, move, nod, offend, pat, play, pour, punish, read, rise, rob, save, scream, see, seek, sell, sew, shiver, shout, sigh, sing, sit, sleep, smash, smile, stand, stop, stride, sue, swim, talk, teach, tell, think, vote, walk, wear

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 16 February 2012
doi: 10.3389/fnhum.2012.00022

The neural correlates of the body-object interaction effect in semantic processing

Ian S. Hargreaves[1]*, Gemma A. Leonard[1], Penny M. Pexman[1], Daniel J. Pittman[3], Paul D. Siakaluk[4] and Bradley G. Goodyear[2,3]

[1] *Department of Psychology, University of Calgary, Calgary, AB, Canada*
[2] *Department of Radiology, University of Calgary, Calgary, AB, Canada*
[3] *Seaman Family MR Research Centre, University of Calgary, Calgary, AB, Canada*
[4] *Department of Psychology, University of Northern British Columbia, Prince George, Canada*

Edited by:
Melvin Yap, National University of Singapore, Singapore

Reviewed by:
Carrie Esopenko, Rotman Research Institute, Canada
Anna M. Woollams, University of Manchester, UK

****Correspondence:***
Ian S. Hargreaves, Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada.
e-mail: ishargre@ucalgary.ca

The semantic richness dimension referred to as body-object interaction (BOI) measures perceptions of the ease with which people can physically interact with words' referents. Previous studies have shown facilitated lexical and semantic processing for words rated high in BOI, e.g., *belt*, than for words rated low in BOI, e.g., *sun.* These BOI effects have been taken as evidence that embodied information is relevant to word recognition (Siakaluk et al., 2008a). However, to date there is no evidence linking BOI manipulations to differences in the utilization of perceptual or sensorimotor areas of the brain. The current study used event-related functional magnetic resonance imaging (fMRI) to examine the neural correlates of BOI in a semantic categorization task (SCT). Sixteen healthy adults participated. Results showed that high BOI words were associated with activation in the left inferior parietal lobule (supramarginal gyrus, BA 40), a sensory association area involved in kinesthetic memory. These results provide evidence that the BOI dimension captures the relative availability of sensorimotor information, and that this contributes to semantic processing.

Keywords: semantic richness, body-object interaction, semantic processing, semantic categorization task, fMRI, sensorimotor

Our ability to efficiently extract information from the world around us is a crucial component of human cognition. The debate concerning how and what information is preserved in memory has been a topic of interest to cognitive scientists since the latter half of the twentieth century. Heavily influenced by functionalism, much of this work has focused on the explanatory power of computational models (Pylyshyn, 1984). This yielded a literature with numerous models of conceptual representation, each providing a specification of the mechanisms by which humans acquire and represent information (Morton, 1969; Collins and Loftus, 1975). Though these models often differed in architectural principles, many authors agreed (at least tacitly) that conceptual representation should be understood as fundamentally amodal and symbolic in nature (Pylyshyn, 1984). Recent work has expanded upon this framework, arguing that a more refined model of semantic processing requires the contributions of both amodal systems and modality-specific systems (Patterson et al., 2007; Dove, 2009). An alternative approach known as embodied cognition goes further still, arguing that most, if not all, of semantic processing is tied to the sensory and motor systems that guide our interactions with the world (Barsalou, 1999; Gibbs, 2006; Pulvermüller, 2010).

Since our histories of sensorimotor interactions help to form conceptual representations, many theories of embodied cognition predict the existence of sensorimotor effects during all kinds of processing. This is the case even for processing that is thought to be of a very abstract character, such as the processing of written language. Wilson (2002) argued that the same mechanisms that are involved in perception and action should also play a role in cognition that is decoupled from the processing of the immediate environment. That is, during "off-line" processing where the immediate environment is merely referenced (e.g., when reading individual words) modality-specific systems will still be recruited in order to "assist in thinking and knowing" (p. 633). These predictions have inspired investigations of sensorimotor effects in basic lexical processes.

Using auditorily presented words in a lexical decision task (LDT: "Is it a real English word?"), Myung et al. (2006) found that participants were faster to verify a target word (e.g., "*baseball*") when it was preceded by a prime that shared manipulation features (e.g., "*grenade*") than when compared to an unrelated prime (e.g., "*leaflet*"). The Myung et al. results suggest that the sensorimotor information in the meanings of auditorialy presented nouns is automatically recruited during a LDT, and that overlap in sensorimotor information between meanings can facilitate word recognition. Using a LDT and a phonological lexical decision task (PLDT: "Does it sound like a real English word?") with visually presented words, Siakaluk and colleagues found significant facilitation for words that were rated high in body-object interaction (BOI), a dimension which measures perceptions of the ease with which a human body can physically interact with a word's referent (Siakaluk et al., 2008a; Tillotson et al., 2008). These effects were observed for a set of high BOI items that were matched with a set of low BOI items on numerous lexical and semantic dimensions such as imageability and concreteness. Thus, the incremental effects of BOI provide evidence that lexical semantics includes information about sensorimotor experiences, and that the relative availability of this information can influence word recognition (Pexman et al., 2002). Using a semantic categorization task (SCT: "is the words' referent easily imageable?") that is thought to focus more on the activation of meaning per se than either the LDT or PLDT, Siakaluk and colleagues observed significant facilitation for words that were rated high on the BOI dimension; that is, faster and more accurate categorization for high BOI words than for low BOI words (Siakaluk et al., 2008b; Bennett et al., 2011; Wellsby et al., 2011). Again, the authors interpreted these significant BOI effects as evidence that sensorimotor information is incorporated in lexical semantics and that this information influences off-line cognition in which the physical environment is merely referenced using language.

Along with BOI effects, other research has demonstrated that participants' metalinguistic judgments about the sensorimotor characteristics of words are able to predict word recognition performance. In a related study by Juhasz and colleagues (Juhasz et al., 2011) participants were directed to not only consider the ease of embodied interaction with the referent of a given word's meaning, but also the degree to which additional sensory experiences (e.g., taste, smell, sight, sound) are evoked by that word. These sensory experience ratings (SERs) were found to account for a significant and unique proportion of variance in LDT reaction times for over 2000 words from the British Lexicon Project (Keuleers et al., 2012). While BOI ratings are assumed to capture sensorimotor experience, the additional variance in LDT that SER ratings accounts for is thought to result from the added contributions of the other senses. However, this explanation relies on an untested assumption that is the focus of the present research. Though Siakaluk and colleagues have consistently found evidence of BOI effects in lexical semantic tasks, questions remain as to what BOI ratings actually capture. Possibilities include proprioceptive or kinesthetic information, such as motor programs for effectively interacting with the environment that are then stored in memory (Gibbs, 2006). However, it is possible that many kinds of information influence participants' BOI ratings, and to date there is no evidence linking BOI effects to activation in areas of the brain dedicated to sensorimotor or kinesthetic memory. The objective of the current study is to test the assumption that BOI ratings capture sensorimotor experience by investigating the neurophysiological correlates of BOI effects during a SCT.

Very few studies have investigated the neural correlates of semantic richness effects in visual word recognition. Using functional magnetic resonance imaging (fMRI), Pexman and colleagues found that words with a greater number of associates (NoA; Nelson et al., 1998) showed less cortical activation than words with fewer associates in an SCT (Pexman et al., 2007). The authors attributed this to the relative efficiency of processing, with low NoA words requiring more processing time, and recruiting more cortical areas, than high NoA words. Though studied by separate labs and in separate tasks (LDT and SCT, respectively) both NoA (Müller et al., 2010) and the number of semantic features (NoF: McRae et al., 2005; Amsel, 2011) displayed unique time courses and cortical topographies relative to other lexical predictors in electroencephalography (EEG) studies. Although it is difficult to extend these findings to the present study in order to make predictions about BOI, the results of these studies clearly demonstrate that semantic richness variables can influence cortical activity during visual word recognition.

Perhaps more useful for forming predictions about BOI effects are the results of numerous studies which have revealed contributions of modality-specific areas of the brain when participants are engaged in off-line processes such as reading (Pulvermüller, 2010; Kiefer and Pulvermüller, 2012). One example is a study by Pulvermüller et al. (2005; see also Hauk and Pulvermüller, 2004). Participants were asked to make lexical decisions about action words that involved face, arm, and leg actions. They used sub-threshold TMS to stimulate areas of the motor cortex, targeting arm and leg areas of the left language-dominant hemisphere. Participants were instructed to respond in the LDT by making brief lip movements in order to avoid confounding TMS stimulation with a manual response. Pulvermüller and colleagues found that targeted TMS stimulation improved the recognition of action words in the LDT. Moreover, this effect was somatotopically mapped, so that, for example, stimulation of arm-areas enhanced processing of arm-related action words compared to leg-related action words. In related work, Desai et al. (2010) used fMRI to examine the cortical activation associated with auditory processing of sentences describing motor actions of the hand/arm (compared to sentences describing visual events or abstract behaviors). Results showed greater activation for these motor action sentences in several sensorimotor regions, including left inferior post-central sulcus and supramarginal gyrus (BA 40). If the BOI variable does indeed measure sensorimotor experience then some of these areas may also be associated with processing of high BOI words in the present study.

METHOD

PARTICIPANTS

The participants were 16 healthy adults, including eight men ($M = 26.50$ years, SD $= 7.15$ years) and eight women ($M = 22.12$ years, SD $= 1.72$ years), all paid for participation. All participants were right-handed, monolingual English speakers, with normal or corrected-to-normal vision. Participants had no history of psychological or developmental disorders, neurological impairments, or any prescription drug use at the time of participation.

MATERIALS

SELECTION OF STIMULI

Stimuli were selected from items in the BOI rating norms acquired by Tillotson et al. (2008). A total of 72 words were selected from the norms, with 36 words rated high in BOI (e.g., *belt*) and 36 words rated low in BOI (e.g., *sun*). Care was taken to ensure that these two sets of items were matched with respect to other lexical and semantic variables that are known to influence behavior and correlated neural activity (see **Table 1**), and these procedures are outlined below. Following McRae and colleagues (McRae et al., 2005), a separate group of 28 participants completed an online ratings task, and were asked to list different types of features, such as physical properties (how it looks), and

Table 1 | Mean characteristics (Standard Deviations in parentheses) for word stimuli.

Word type	BOI	Length	NoF	Familiarity	Concreteness	Imageability	Print frequency	CD	OLD20	PLD20
High BOI	5.60	4.20	8.47	542.48	560.25	560.25	3.24	2.95	1.38	1.21
	(0.47)	(0.78)	(2.44)	(49.00)	(50.00)	(46.00)	(0.65)	(0.58)	(0.32)	(0.26)
Low BOI	3.30	4.20	8.05	531.68	550.00	556.50	3.26	2.97	1.39	1.10
	(0.59)	(0.81)	(3.00)	(46.58)	(48.00)	(47.00)	(0.80)	(0.50)	(0.27)	(0.25)
p-value	<0.001	0.88	0.61	0.86	0.55	0.59	0.90	0.65	0.87	0.73

Note: p-values reflect difference test between high and low BOI word types; BOI = rated body-object interaction [Tillotson et al. (2008)]; Length = length in letters; NOF = Number of features [McRae et al. (2005)]; Familiarity = rated familiarity [MRC Database, Coltheart (1981)]; Concreteness = rated concreteness [MRC Database, Coltheart (1981)]; Imageability = rated imageability [MRC Database, Coltheart (1981)]; Print frequency = log10 frequency of occurrence in print [Brysbaert and New (2009)]; CD = log10 contextual diversity [Brysbaert and New (2009)]; OLD20 = orthographic Levenshtein distance 20 [Yarkoni et al. (2008)]; PLD20 = phonological Levenshtein distance 20 [Yarkoni et al. (2008)].

functional properties (what it is used for) for each target concept. For each concept, the features listed were recorded along with the number of participants who listed each feature. Results of the feature listing task showed that the high BOI and low BOI word lists did not differ on number of features per concept. In addition, the two word sets were matched for length, printed frequency, contextual diversity (Brysbaert and New, 2009), subjective familiarity (Balota et al., 2001), the mean Levenshtein distance of a word to its 20 closest orthographic and phonological neighbors (Yarkoni et al., 2008), and importantly, concreteness and imageability (Cortese and Fugett, 2004). The descriptive statistics for all the items are presented in **Table 1**. Finally, an additional 60 less imageable nouns (e.g., *rate*) were selected for the "no-go" trials, yielding a total of 132 trials. A slight imbalance in the number of critical (72) and distractor (60) trials was created in order to increase the number of critical trials included in the analysis without increasing the amount of time spent in-scanner. To be clear, the critical items were high and low BOI words for which subjectively rated imageability and concreteness had been controlled (**Table 1**), and thus any observed effects of BOI in this SCT can be interpreted as incremental to those of imageability or concreteness.

PROCEDURE

The study was conducted at the Seaman Family MR Research Center at the Foothills Hospital, located in Calgary, Alberta. This study was reviewed and approved by the University of Calgary Research Ethics Board. Participants were informed of any risks associated with participating, and written consent was obtained from all participants prior to partaking in the study.

A trial was initiated by a fixation marker that appeared at the center of the computer display for 1000 msec, and was then replaced by a word. Stimuli were presented for 2500 msec with a randomized interval of 4000 msec $\pm$ 2000 msec. A variable inter-trial-interval was used to increase the detectability of the hemodynamic responses to trials (Birn et al., 2002). All stimuli were presented in a randomized order in a single block lasting 17 min. Previous investigations of BOI effects in behavior have demonstrated a significant influence of BOI during a SCT when participants were asked to decide whether a word was easily imageable or not (Siakaluk et al., 2008b). These effects of BOI on categorization performance were quite robust, and have been observed using both manual and verbal responses during a go/no-go SCT using the imageability decision (Wellsby et al., 2011). As this was the first investigation into the brain-based correlates of BOI effects, we followed Wellsby and colleagues' procedure, adopting a go/no-go SCT using the imageability decision category. This procedure was adopted in order to facilitate the interpretation of the imaging results by using a SCT that has already demonstrated a sizable (61 msec) behavioral effect of BOI. In addition, the imageability decision is sufficiently broad that a number of both high and low BOI items could be equally typical of the decision category, while at the same time requiring that participants engage in semantic processing in order to respond accurately. The participants' task was to decide whether each word's referent was easily imageable and to respond as quickly and as accurately as possible. Participants were instructed to respond only to the words that were easily imageable ("go" response) by pressing the button and were further instructed to refrain from making a button response to the words that were less imageable ("no-go" response). After a button press, the word would appear underlined, so as to indicate that a response had been made. Each participant first completed six practice trials, outside of the magnet room, consisting of four words that were imageable and two that were less imageable. All practice stimuli were similar in normative frequency to the experimental stimuli.

ACQUISITION OF BEHAVIORAL DATA

All stimuli were presented to participants using a rear-mounted LED projector display system (Avotec, Inc., Stuart, FL). The sequence of trials was presented using Presentation software (Neurobehavioral Systems, Albany, CA), running on a computer located outside of the magnet room. Participants' responses were recorded using a MR-compatible Lumina response pad (Cedrus Corporation, San Pedro, CA).

ACQUISITION OF fMRI DATA

Images were acquired using a 3-Tesla General Electric MR scanner, equipped with an 8-channel phased array head coil (Signa Excite; GE Healthcare, Waukesha, WI). The MR sequence for functional imaging was a single-shot gradient-recalled echo planar imaging (EPI) T2*-weighted sequence, with whole head coverage (64×64 matrix, zero-filled to 128×128, FOV = 24 cm, TE = 30 msec, TR = 2000 msec, flip angle = 70, 31 oblique/axial slices, 4 mm thick). High-resolution T1-weighted images (0.94 × 0.94 × 2.00 mm) were collected using a 3D inversion recovery-prepped anatomical MRI sequence to anatomically register the functional data.

Image analyses were carried out using FEAT (FMRI Expert Analysis Tool) Version 5.98, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Images were corrected for head motion during post-processing using the intra-modal motion correction tool MCFLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002). Prior to analysis, image data were subjected to high-pass temporal filtering (Gaussian-weighted LSF straight line fitting, with sigma = 25.0 s). The following pre-statistics processing was also applied: slice-timing correction using Fourier-space time-series phase-shifting; non-brain removal using BET (Smith, 2002); spatial smoothing using a Gaussian kernel of FWHM 6 mm; and mean-based intensity normalization of all volumes by the same factor. For all participants, time-series statistical analyses for each stimulus category (high and low BOI trials, no-go trials, and error trials) were carried out using FMRIB's Improved Linear Model (FILM) with local autocorrelation correction (Woolrich et al., 2001). Registration of high-resolution images to the MNI brain was carried out using FLIRT (Jenkinson and Smith, 2001; Jenkinson et al., 2002).

Higher-level contrasts of planned comparisons of the critical stimuli (high vs. low BOI words) across participants were carried out using FLAME (FMRIB's Local Analysis of Mixed Effects) stage 1 only (i.e., without the final MCMC-based stage) (Beckmann et al., 2003; Woolrich et al., 2004). Z-statistic images for these analyses were generated using a random effects model and a statistical threshold of $Z = 2.3$, and a cluster size threshold of at least 77 contiguous voxels corresponding to a corrected p-value of 0.05 as determined by Monte Carlo simulations using AlphaSim (http://afni.nihm.nih.gov/afni/doc/manual/AlphaSim). These simulations provide an estimate, for a given smoothness, of the cluster volume necessary to exceed a certain confidence in a cluster.

RESULTS

BEHAVIORAL RESULTS

Words with less than 70% accuracy (3 high BOI and 7 low BOI, identified in the Appendix) were excluded from subsequent analysis[1]. No trials were excluded on the basis of response latency, as all responses fell between 250 msec and 2500 msec. Responses to the high BOI words ($M = 930$ msec, SD = 154 msec) were significantly faster than responses to the low BOI words ($M =$ 1083 msec, SD = 153 msec), and this BOI effect was significant by subjects and by items, $t_1(15) = 6.62$, $p < 0.001$; $t_2(60) = 4.66$, $p < 0.001$. There was also a BOI effect in the accuracy data, as responses to the high BOI words ($M = 98\%$, SD = 4%) were significantly more accurate than responses to the low BOI words ($M = 93\%$, SD = 7%) both by subjects and by items, $t_1(15) =$ 5.75, $p < 0.001$; $t_2(60) = 3.37$, $p = 0.001$. These findings replicate the facilitatory BOI effects reported by Bennett et al. (2011), Siakaluk et al. (2008b), and Wellsby et al. (2011).

IMAGING RESULTS

The results of the planned contrasts of high BOI and low BOI words are displayed in **Figure 1** and all significant regions of activation, z-scores, and corresponding Talairach coordinates from these contrasts, are presented in **Table 2**. Several areas were significantly more active during categorization of low BOI words compared to high BOI words. These areas are part of an inhibitory control network often observed in go/no-go paradigms (e.g., Nakata et al., 2008; Simmonds et al., 2008), and comprise the right superior frontal gyrus (BA 8) including the pre-supplementary motor area (BA 6), the right middle frontal gyrus (BA 9), and the right inferior frontal gyrus (BA 45). In the reverse contrast, one area was significantly more active in the processing of high BOI words compared to low BOI words. That is, we observed greater activation in the left inferior parietal lobule (supramarginal gyrus, SMG, BA 40), a sensory association area involved in kinesthetic memory (Grèzes and Decety, 2001; Péran et al., 2010).

DISCUSSION

The present study was designed to investigate the neural correlates of BOI effects in semantic categorization. Analyses showed that a number of cortical areas were more active for low BOI words than for high BOI words, and our interpretation of this activity is informed by our use of the go/no-go task. That is, areas that showed more activity for low BOI words compared to high BOI words include a network thought to underlie response inhibition (Nakata et al., 2008). This relative utilization of inhibitory control mechanisms suggests that, compared to high BOI words, low BOI words don't contribute as much positive information in favor of a "go" response. Hence, more activity is observed in the inhibitory control network when participants categorize low BOI words. This is consistent with past interpretations of BOI

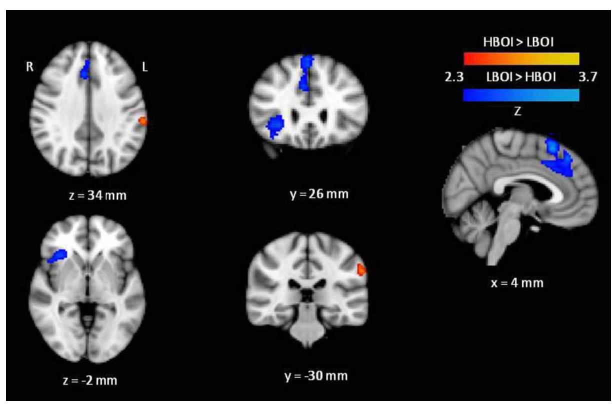


FIGURE 1 | Cortical activation maps displaying results of the high BOI (HBOI) versus low BOI (LBOI) contrasts. Regions showing greater activation for high BOI compared to low BOI are shown in orange. Regions showing greater activation for low BOI compared to high BOI are shown in Blue.

[1]The adoption of these exclusion criteria did not create a significant imbalance between high and low BOI lists in terms of the control variables reported in **Table 1** (all p-values > 0.05) and there remained a significant difference along the BOI dimension between high ($M = 5.61$, SD = 0.46) and low BOI ($M =$ 3.37, SD = 0.58) lists, $t(60) = 16.75$, $p < 0.001$.

Table 2 | Areas of significant activation in contrasts between word types.

Contrast	Region of activation	BA	*Z* Score	Talaraich coordinates *x, y, z*
High BOI > Low BOI	Left Supramarginal Gyrus	40	3.06	–62, –30, 34
Low BOI > High BOI	Right Superior Frontal Gyrus (pre-SMA)	6	3.11	4, 22, 58
	Right Superior Frontal Gyrus	8	2.72	4, 38, 46
	Right Middle Frontal Gyrus	9	2.42	48, 16, 26
	Right Inferior Frontal Gyrus	45	2.49	36, 28, –2

Note: Brodmann's areas (BA) should be considered estimates only.

effects on semantic categorization (Siakaluk et al., 2008b; Bennett et al., 2011), which have argued that the observation of behavioral facilitation for high BOI words stems from the relative availability of semantic information for the decision. In the present version of the go/no-go task, participants must withhold a response for abstract words. Since high BOI words are relatively richer in terms of their semantics, they contribute relatively more information to the decision than do low BOI words. In contrast, the observed cortical activation suggests that for low BOI word participants show a tendency to withhold their response, despite successfully responding "go" for these words. Again, this is consistent with the supposition that low BOI words are contributing relatively less information to the decision.

As reviewed above, the BOI effect is thought to stem from the relative availability of sensorimotor information for the go/no-go decision. We tested this assumption by contrasting areas that were significantly more active for high BOI words than low BOI words. Results of this contrast showed activation in the left inferior parietal lobule (SMG, BA 40). This area has been implicated in perception and planning of goal-oriented hand-object interactions (Russ et al., 2003; Naito and Ehrsson, 2006; Tunik et al., 2008). In addition, data from clinical populations suggests that the parietal cortex is a central area for the storage and subsequent access of motor information. Lesions to the SMG have been associated with ideomotor apraxia (Haaland et al., 2000), which is characterized by an inability to correctly plan and execute motor programs when given a verbal command. This can include the inability to correctly pantomime the use of an object (e.g., "Pantomime combing your hair.") or an inability to use objects properly in real-life situations (Wheaton and Hallett, 2007). These apraxic deficits are most commonly observed for actions that are directed toward objects or tools. Importantly, in order to qualify as ideomotor apraxia, object-knowledge and the ability to correctly recognize objects must be preserved.

During the BOI ratings task (Tillotson et al., 2008), participants were given explicit instructions to rate each word in terms of the ease with which the human body can physically interact with the word's referent, and to try to ignore other related-factors, such as how easily it can be experienced by the senses (e.g., vision, taste, etc.). As a result, the assumption is that these ratings capture the relative degree of sensorimotor experience that participants have with the object to which the word refers. It was unclear, however, to what extent these ratings actually captured differences in sensorimotor information. The present finding, that words rated high in BOI recruit areas of the brain (during an off-line visual word recognition task) that play an important role in kinesthetic memory, specifically kinesthetic memory that is involved in the correct performance of verbally cued complex actions for objects, supports the assumption that BOI ratings capture differences in sensorimotor experience.

Though the relationship between SMG activation and BOI is easily interpretable, the specificity of the observed cortical activation raises an interesting question. As reviewed in the introduction, the studies of Pulvermüller and colleagues suggest that somatotopically mapped areas of the motor cortex play a functional role in the processing of action words (Hauk and Pulvermüller, 2004; Pulvermüller et al., 2005). Indeed, a large distributed network of modality-specific areas is implicated in semantic processing (Patterson et al., 2007; Kiefer and Pulvermüller, 2012). Why, then, does a manipulation of the perceived ease of embodied interaction with a words' referent selectively recruit the SMG? Certainly, the demands of the go/no-go task could limit the contributions of other cortical areas to processing (Hargreaves et al., 2012). However, it is also worth noting that the current sets of stimuli are carefully balanced on numerous lexical and semantic dimensions (**Table 1**). Thus, although a distributed network of modality-specific areas may contribute to the construction of word meaning, our controlled manipulation of relative BOI may render our analysis insensitive to many of these contributions.

By manipulating BOI, we observed that the relative availability of sensorimotor information influenced activity in an area of the brain that is involved in complex motor processing. As such, our results are similar to those of other studies showing that sensorimotor and perceptual systems contribute to off-line processing such as reading (e.g., Desai et al., 2010). Studies have shown that language processing can rapidly recruit areas of the brain dedicated to perception and action, suggesting an immediate role for this information in the construction of meaning (Pulvermüller, 2010). However, there is always the possibility that this information is ancillary, and does not directly contribute to core semantic processes (Mahon and Caramazza, 2008). Just as there are many models of semantic memory, there is a great deal of variability in what researchers consider to be semantic. A broad definition of semantics as "world knowledge" would easily integrate the current findings, with BOI capturing the contribution of sensorimotor information to the processing of concepts. This broad definition has been utilized by our group to interpret the behavioral consequences of BOI in past studies (Siakaluk et al., 2008a; Bennett et al., 2011), and is also well represented by recent developments in cognitive neuroscience. For example, a framework like that proposed by Kiefer and Pulvermüller (2012) features concepts that are flexible, and that have representations that are distributed across numerous modality-specific informational dimensions. However, a more narrow definition of semantics that construes motor contributions as auxiliary to core semantic processing would not accommodate motor processes as constitutive of meaning, only that they interface with the conceptual system at some point during processing (Mahon and Caramazza, 2008). Note that neither side would disagree on the current data, only on the extent to which we can claim that sensorimotor information contributes to the construction of meaning.

The present results provide valuable new insight about the nature of BOI effects in visual word recognition. This study was the first test of the brain-based consequences of this variable. The results showed that higher levels of BOI are associated with activity in an area of the brain involved in kinesthetic memory, supporting the assumption of Siakaluk and colleagues that variability along the BOI dimension captures variability in the availability of sensorimotor information (2008b). Our results build upon a literature documenting effects of sensorimotor experience in reading (Wilson, 2002; Siakaluk et al., 2008a), a form of off-line processing that is somewhat removed from physical interaction with the environment and was originally conceived as the systematic manipulation of abstract symbols that are fundamentally amodal in character (Pylyshyn, 1984). With additional studies, researchers may better understand whether the contribution of sensorimotor information is indicative of a dynamic semantic system that utilizes multiple, modality-specific sources (Kiefer and Pulvermüller, 2012), or is best conceived as the manipulation of abstract structures, that interface with modality-specific processes downstream (Mahon and Caramazza, 2008). Regardless of which view one subscribes to, the present results serve to clarify and reinforce the contributions of sensorimotor experience to lexical-semantic processing.

REFERENCES

Amsel, B. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia* 49, 970–983.

Balota, D., Pilotti, M., and Cortese, M. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Mem. Cognit.* 29, 639–647.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Beckmann, C., Jenkinson, M., and Smith, S. M. (2003). General multilevel linear modeling for group analysis in fMRI. *Neuroimage* 20, 1052–1063.

Bennett, S. D. R., Burnett, A. N., Siakaluk, P. D., and Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599 multisyllabic words. *Behav. Res. Methods* 43, 1100–1109.

Birn, R. M., Cox, R. W., and Bandettini, P. A. (2002). Detection versus estimation in event-related fMRI: choosing the optimal stimulus timing. *Neuroimage* 15, 252–264.

Brysbaert, M., and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods Instrum. Comput.* 41, 977–990.

Coltheart, M. (1981). MRC psycholinguistic database. *Q. J. Exp. Psychol.* 33A, 497–505.

Collins, A., and Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428.

Cortese, M., and Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behav. Res. Methods Instrum. Comput.* 36, 384–387.

Desai, R. H., Binder, J. R., Conant, L. L., and Seidenberg, M. S. (2010). Activation of sensory-motor areas in sentence comprehension. *Cereb. Cortex* 20, 468–478.

Dove, G. (2009). Beyond perceptual symbols: a call for representational pluralism. *Cognition* 110, 412–431.

Gibbs, R. W. (2006). *Embodiment and Cognitive Science*. New York, NY: Cambridge University Press.

Grèzes, J., and Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12, 1–19.

Haaland, K., Harrington, D., and Knight, R. (2000). Neural representations of skilled movement. *Brain* 123, 2306–2313.

Hargreaves, I. S., White, M., Pexman, P., Pittman, D., and Goodyear, B. (2012). The question shapes the answer: the neural correlates of task differences reveal dynamic semantic processing. *Brain Lang.* 120, 73–78.

Hauk, O., and Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Hum. Brain Mapp.* 21, 191–201.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841.

Jenkinson, M., and Smith, S. M. (2001). A global optimization method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156.

Juhasz, B., Yap, M., Dicke, J., Taylor, S., and Gullick, M. (2011). Tangible words are recognized faster: the grounding of meaning in sensory and perceptual systems. *Q. J. Exp. Psychol.* 64, 1683–1691.

Keuleers, E., Lacey, P., Rastle, K., and Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behav. Res. Methods* [Epub ahead of Print].

Kiefer, M., and Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex* [Epub ahead of Print].

Mahon, B., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods* 37, 547–559.

Morton, J. (1969). Interaction of information in word recognition. *Psychol. Rev.* 76, 165–178.

Müller, O., Duñabeitia, J., and Carreiras, M. (2010). Orthographic and associative neighbourhood density effects: what is shared, what is different? *Psychophysiology* 47, 455–466.

Myung, J., Blumstein, S., and Sedivy, J. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition* 98, 223–243.

Naito, E., and Ehrsson, H. H. (2006). Somatic sensation of hand-object interactive movement is associated with activity in the left inferior parietal cortex. *J. Neurosci.* 26, 3783–3790.

Nakata, H., Sakamoto, K., Ferretti, A., Gianni Perrucci, M., Del Gratta, C., Kakigi, R., and Luca Romani, G. (2008). Somato-motor inhibitory processing in humans: an event-related functional MRI study. *Neuroimage* 39, 1858–1866.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). *The University of South Florida Word Association, Rhyme, and Word Fragment Norms* [Online]. Available: http://www.usf.edu/FreeAssociation/

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987.

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., and Goodyear, B. G. (2007). The neural consequences of semantic richness: when more comes to mind, less activation is observed. *Psychol. Sci.* 18, 401–406.

Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number of features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.

Péran, P., Démonet, J. F., Cherubini, A., Carbebat, D., Caltagirone, C., and Sabatini, U. (2010). Mental representations of action: the neural correlates of the verbal and motor components. *Brain Res.* 1328, 89–103.

Pulvermüller, F. (2010). Brain-language research: where is the progress? *Biolinguistics* 4, 255–288.

Pulvermüller, F., Hauk, O., Nikulin, V., and Ilmoniemi, R. J. (2005). Functional interaction of language and action: a TMS study. *Eur. J. Neurosci.* 21, 793–797.

Pylyshyn, Z. (1984). *Computation and Cognition: Towards a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.

Russ, M. O., Mack, W., Grama, C. R., Lanfermann, H., and Knopf, M. (2003). Enactment effect in memory: evidence concerning the function of the supramarginal gyrus. *Exp. Brain Res.* 149, 497–504.

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., and Sears, C. R. (2008a). Evidence for the activation of sensorimotor information during visual word recognition: the body-object interaction effect. *Cognition* 106, 433–443.

Siakaluk, P. D., Pexman, P. M., Sears, C. R., Wilson, K., Locheed, K., and Owen, W. J. (2008b). The benefits of sensorimotor knowledge: body-object interaction facilitates semantic processing. *Cogn. Sci.* 32, 591–506.

Simmonds, D., Pekar, J., and Mostofsky, S. (2008). Meta-analysis of go/no-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia* 46, 224–232.

Smith, S. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.

Tillotson, S. M., Siakaluk, P. D., and Pexman, P. M. (2008). Body-Object interaction ratings for 1,618 monosyllabic nouns. *Behav. Res. Methods* 40, 1075–1078.

Tunik, E., Lo, O. Y., and Adamovich, S. V. (2008). Transcranial magnetic stimulation to the frontal operculum and supramarginal gyrus disrupts planning of outcome-based hand-object interactions. *J. Neurosci.* 28, 14422–14427.

Wellsby, M., Siakaluk, P., Owen, W., and Pexman, P. (2011). Embodied semantic processing: the body-object interaction effect in a non-manual task. *Lang. Cogn.* 3, 1–14.

Wheaton, L., and Hallett, M. (2007). Ideomotor apraxia: a review. *J. Neurol. Sci.* 260, 1–10.

Wilson, M. (2002). Six views on embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2004). Multi-level linear modeling for fMRI group analysis using Bayesian inference. *Neuroimage* 21, 1732–1747.

Woolrich, M. W., Ripley, B. D., Brady, J. M., and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage* 14, 1370–1386.

Yarkoni, T., Balota, D., and Yap, M. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychon. Bull. Rev.* 15, 971–979.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 January 2012; paper pending published: 24 January 2012; accepted: 05 February 2012; published online: 16 February 2012.

Citation: Hargreaves IS, Leonard GA, Pexman PM, Pittman DJ, Siakaluk PD and Goodyear BG (2012) The neural correlates of the body-object interaction effect in semantic processing. Front. Hum. Neurosci. ***6****:22. doi: 10.3389/fnhum.2012.00022*

APPENDIX

HIGH BOI

Belt, boot, cage, card, cart, child, cord, couch, drill, feet, friend*, gate, hat, hook, lake, lock, mail, man, mat, mate*, neck, priest, purse, room, seat, silk, stair, string, suit, sword, thing*, tool, toy, tube, vest, wheel.

LOW BOI

Ash, back, band, bay, birch, brain, brass, case, coast, frost, game*, gang, heart, jail, king, knight, lane, lint, loot*, lung, place*, prince, pump, roof, shop, slit, song*, spot, stripe, sun, trail, tar, trip*, war*, well*, witch.

*Items removed from analyses due to high error rates.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 20 April 2012
doi: 10.3389/fnhum.2012.00084

A neuroanatomical examination of embodied cognition: semantic generation to action-related stimuli

Carrie Esopenko[1]*, Layla Gould[2], Jacqueline Cummine[3], Gordon E. Sarty[2], Naila Kuhlmann[2] and Ron Borowsky[2]*

[1] *Baycrest Centre, Rotman Research Institute, University of Toronto, ON, Canada*
[2] *Department of Psychology, College of Arts and Science, Cognitive Neuroscience of Language Lab, University of Saskatchewan, Saskatoon, SK, Canada*
[3] *Department of Speech Pathology and Audiology, University of Alberta, Edmonton, AB, Canada*

Edited by:
Penny M. Pexman, University of Calgary, Canada

Reviewed by:
Diane Pecher, Erasmus University Rotterdam, Netherlands
Daniel Bub, University of Victoria, Canada

****Correspondence:***
Ron Borowsky, Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK S7N 5A5, Canada.
e-mail: ron.borowsky@usask.ca

Carrie Esopenko, Rotman Research Institute, Baycrest Centre, 3560 Bathurst St. Toronto, ON M6A 2E1, Canada.
e-mail: cesopenko@rotman-baycrest.on.ca

The theory of embodied cognition postulates that the brain represents semantic knowledge as a function of the interaction between the body and the environment. The goal of our research was to provide a neuroanatomical examination of embodied cognition using action-related pictures and words. We used functional magnetic resonance imaging (fMRI) to examine whether there were shared and/or unique regions of activation between an ecologically valid semantic generation task and a motor task in the parietal-frontocentral network (PFN), as a function of stimulus format (pictures versus words) for two stimulus types (hand and foot). Unlike other methods for neuroimaging analyses involving subtractive logic or conjoint analyses, this method first isolates shared and unique regions of activation within-participants before generating an averaged map. The results demonstrated shared activation between the semantic generation and motor tasks, which was organized somatotopically in the PFN, as well as unique activation for the semantic generation tasks in proximity to the hand or foot motor cortex. We also found unique and shared regions of activation in the PFN as a function of stimulus format (pictures versus words). These results further elucidate embodied cognition in that they show that brain regions activated during actual motor movements were also activated when an individual verbally generates action-related semantic information. Disembodied cognition theories and limitations are also discussed.

Keywords: embodied cognition, fMRI, semantic generation, pictures, words, action-related

Determining how the interaction between the body and the environment influences the evolution, development, organization, and processing of the human brain is important in the understanding of how conceptual information is represented (Barsalou, 1999; Wilson, 2002; Gibbs, 2006; Siakaluk et al., 2008a). According to Barsalou (1999) Perceptual Symbols Systems theory, there is a strong relationship between an individual's perceptual experiences and his or her conceptual representations. Specifically, the perceptual symbols or representations of an experience are encoded in the brain, these representations form a conceptual simulation of that experience, and when an individual retrieves information about a concept, the perceptual symbols (or experiences) associated with the concept are simulated. As such, conceptual information is grounded in perceptual, or sensorimotor, experiences that are simulated when re-enacting interactions with stimuli (see also Barsalou, 2008, 2009). These notions are the theoretical underpinnings of embodied cognition, which proposes that cognitive processes are embedded in sensorimotor processing (Wilson, 2002; Siakaluk et al., 2008b). Accordingly, embodied cognition theorists suggest that cognition is bodily based, in that the mind is used to guide action, and that the brain developed as a function of interaction with the environment to facilitate sensory and motor processing (Wilson, 2002; Gibbs, 2006). As such, the theory of embodied cognition would suggest that conceptual information is grounded in sensorimotor processes, and thus sensorimotor regions involved in encoding conceptual knowledge should be active during conceptual (i.e., semantic) processing.

NEUROIMAGING EVIDENCE OF EMBODIMENT

A recent goal of cognitive neuroscience has been to determine whether the brain regions that control actions are also involved when responding to action-related language (for a review see Willems and Hagoort, 2007). That is, are motor regions active during responding to semantic tasks through using action-related language, even when the task involves no motor movement or actions? If sensorimotor regions are active when using action-related language during semantic tasks, this could be taken as neuroanatomical support for embodied cognition theories. In the action-semantics and embodied cognition research literature, activation of the parietal-frontocentral network (PFN; including the supramarginal gyrus, inferior and superior parietal lobule, and sensory cortex of the parietal lobes, and the supplementary motor area (SMA), and the premotor regions of the frontal lobes) when responding to action-related language, is typically referred to as *somatotopic-semantics*. As such, evidence of somatotopic-semantics supports the idea that action-related conceptual knowledge is grounded in action and perceptual systems (Pulvermuller, 2005; Barsalou, 2008; Boulenger et al., 2009).

Recent neuroimaging studies have shown that sensorimotor regions are activated somatotopically during the processing of action-related language, much like Penfield's map of the sensory and motor homunculus (Penfield and Rasmussen, 1950), and this has been taken as neuroanatomical evidence of embodied cognition (Hauk et al., 2004; Tettamanti et al., 2005; Esopenko et al., 2008; Boulenger et al., 2009; Raposo et al., 2009). Such research has shown that action-related language activates the PFN and is organized somatotopically dependent upon the body part the object/action represents (Hauk et al., 2004; Pulvermuller, 2005; Tettamanti et al., 2005; Esopenko et al., 2008; Raposo et al., 2009). For example, when participants covertly read action words (Hauk et al., 2004), listen to action-related sentences referring to specific body parts (mouth, arm, leg; Tettamanti et al., 2005), or overtly generate a response to how they would use a hand- or foot-related object presented in word format (Esopenko et al., 2008), regions proximal to where the body part is represented on the motor cortex are activated. Aziz-Zadeh et al. (2006) found that regions activated by observing actions overlapped with regions activated by reading phrases depicting actions in a somatotopic arrangement. As such, the abovementioned research suggests that the PFN is accessed during processing of action-related language in a somatotopic fashion, or in other words, the semantic representations for action-related language are embodied. However, it is important to note that there are some criticisms to embodied cognition theories in the literature. The major criticism is that the motor system is not required for the processing of action-related stimuli, and if activation does occur in the motor system, it happens after the semantic analysis of the stimulus has occurred (Caramazza et al., 1990; Mahon and Caramazza, 2005, 2008). In other words, according to this view, the motor system is not required for processing action-related stimuli, but is an automatic by-product of processing such stimuli. We will return to this issue in the Discussion.

PROCESSING OF ACTION-RELATED PICTURE AND WORD STIMULI

Previous neuroimaging research examining embodied cognition has not directly compared activation in the regions that have been shown to process embodied information (i.e., in the PFN), using the same stimuli in both picture and word format. Previous patient research has shown that patients can present with deficits in recognition dependent upon stimulus format (e.g., Lhermitte and Beauvois, 1973; Bub et al., 1988; Lambon Ralph and Howard, 2000). Behavioral research has also shown that pictures and words have differential access to semantic action-related knowledge. Specifically, Thompson-Schill et al. (2006) have suggested that pictorial stimuli contain form information and thus have privileged access to manipulation knowledge compared to word stimuli. For example, Chainey and Humphreys (2002) found that participants are faster at making action decisions to picture stimuli compared to word stimuli, which is suggested to occur because stored associations for actions are more easily accessed by the visual properties of the object. In addition, Saffran et al. (2003) have shown that a significantly greater number of verbs are produced for pictures compared to words. The authors suggest that this could occur because pictures provide the affordance to how the object can be used, thus making it easier to produce a verb representing the object use. Such research suggests that although pictures and words have access to general semantic knowledge, there is evidence that there are differences in the information that pictures and words activate. Thus, it is important to examine whether pictures and words activate unique, as well as common, brain regions that process action-related stimuli.

OUR RESEARCH

Our research examined whether there were differences in the neuroanatomical processing of embodied, or action-related, stimuli presented in picture and word format. To examine this, participants completed either the hand or foot variant of the semantic generation task from Esopenko et al. (2008, 2011), with the same stimuli presented in both picture and word formats to directly compare whether there are differences in processing in the PFN dependent upon picture versus word format. Furthermore, we sought to examine whether the semantic generation and motor localization tasks showed a shared network of activation in the PFN. We compared a word semantic generation task, a picture semantic generation task, and a motor localization task to examine the following hypotheses: (1) to the extent that conceptual knowledge is grounded in sensorimotor processing (i.e., embodied), the semantic generation tasks should activate the PFN in a somatotopic fashion (i.e., in proximity to the hand and foot sensory and motor cortices); (2) to the extent that this network in the PFN is common to both picture and word format, we should see similar shared activation maps between the semantic generation and motor localization tasks for both pictures and words; and (3) to the extent that pictures and words have differential access to action-related knowledge, there should also be unique regions of activation for pictures versus words representing the embodied action-related knowledge in the PFN.

To examine the extent of unique and overlapping activation in the PFN for the processing of action-related stimuli, we use unique and shared activation maps that were developed in our lab (Borowsky et al., 2005a,b, 2006, 2007). These shared maps allow one to determine, within-participants, what activation is common between two tasks, whereas unique activation maps allow one to determine what regions are uniquely activated for each task. This differs from the traditional examination of what is unique to each task using subtraction activation maps, whereby traditional subtraction maps show unique activation based on what task activation is of highest intensity when two tasks are pitted against each other. As such, the unique and shared maps used in the fMRI analysis will allow for an additional perspective on what is unique to each of two processes (e.g., motor vs. semantic; picture vs. word), and also, what is shared between the two processes. Given that the shared/unique maps are computed within-participants and then averaged for the final maps, the final shared map is mathematically independent (not mathematically exclusive) of the final unique map.

METHODS

PARTICIPANTS

University undergraduate students ($N = 16$; mean age $= 23$; all right-handed) with normal or corrected to normal vision participated in this experiment. The research was approved by the University of Saskatchewan Behavioral Sciences Ethics Committee.

STIMULI AND PROCEDURE

In the following experiment, participants completed a motor localization task, and a picture and word semantic generation task. Participants completed both the picture and word semantic generation tasks with either hand or foot stimuli (i.e., eight participants in the foot condition and eight participants in the hand condition). The motor localization task was used to determine the location of hand and foot motor cortex. To allow comparisons to the tasks described below, a visual cue was given on each trial, such that in the hand condition the word "Hand," and in the foot condition the word "Foot," was presented on the screen and participants were instructed to move the body parts that the word represented while it was on the screen. For the hand condition, movement involved sequential bimanual finger-to-thumb movements. For the foot condition, movement involved bimanual foot-pedaling motions. The order of these two motor tasks was counterbalanced across participants. For the foot-pedaling motions, participants were instructed to only move their feet and no other part of their body. By using a large-angled piece of foam under the knees, we were able to ensure that the foot-pedaling condition did not create any motion in the upper body.

For the semantic generation tasks, the stimuli consisted of visually presented pictures and words referring to objects that are typically used by the hand (e.g., stapler) or the foot (e.g., soccerball). There were 50 objects in each of the picture and word conditions (**Appendix A**). The same objects were presented in both the picture and word conditions. Although we are not comparing hand and foot conditions to each other, we nevertheless matched the hand and foot stimuli as closely as possible on length [$t(51) = 1.157$, $p = 0.253$] and subtitle word frequency (SUBTLEX frequency per million words) [$t(51) = -1.173$, $p = 0.246$] using the norms from the English Lexicon Project (Balota et al., 2007). However, some words could not be matched given that some words were not included in the database. Order of presentation format (picture/word) and stimulus type (hand/foot) was counterbalanced across participants. Participants were presented with five blocks of pictures or words (with five words/pictures in each block) referring to objects that are primarily used by the hand or the foot and were instructed to quickly describe how they would physically interact with the object during a gap in image acquisition (i.e., using a sparse-sampling image acquisition method). This paradigm allows the participant to report their own conceptual knowledge about the objects, as opposed to judging whether they agree with some pre-determined categorization of the objects. The gap allowed the experimenter to listen to each response to ensure that the participant provided a response that was appropriate for the task (e.g., Borowsky et al., 2005a, 2006, 2007; Esopenko et al., 2008). An example of a hand response is (e.g., for *pen*) "write with it," and for a foot response is (e.g., for *soccerball*) "kick it."

IMAGING AND IMAGE ANALYSIS

The imaging was conducted using a 1.5 T Siemens Symphony (Erlanger, Germany) magnetic resonance imager. For both the motor and semantic generation tasks, 55 image volumes were obtained, with each image volume consisting of 12 axial slice single-shot fat-saturated echo-planar images (EPI); $T_R = 3300$ ms, with a 1650 ms gap of no image acquisition at the end of the T_R, $T_E = 55$ ms, 64×64 acquisition matrix, 128×128 reconstruction matrix. Each slice was 8 mm thick with a 2 mm thick interslice gap and was acquired in an interleaved sequence (e.g., slices 1–3-5–2–4 etc.) to reduce partial volume crosstalk in the slice dimension. For all tasks, the first five image volumes were used to achieve a steady state and were discarded prior to analysis. The remaining volumes were organized into five blocks of 10 volumes each for a total of 50 image volumes. Each block consisted of five image volumes collected during the presentation of, and response to, the stimuli, followed by five image volumes collected during rest. A computer running E-Prime software (Psychology Software Tools, Inc., Pittsburgh, PA) was used to trigger each image acquisition in synchrony with the presentation of visual stimuli. The stimuli were presented using a data projector (interfaced with the E-prime computer) and a back-projection screen that was visible to the participant through a mirror attached to the head coil. In order to capture a full-cortex volume of images for each participant, either the third or fourth inferior-most slice was centered on the posterior commissure, depending on the superior-inferior distance between the posterior commissure and the top of the brain for each participant. T_1-weighted high-resolution spin-echo anatomical images ($T_R = 400$ ms, $T_E = 12$ ms, 256×256 acquisition matrix, 8 mm slice thickness with 2 mm between slices) were acquired in axial, sagittal, and coronal planes. The position of the twelve T_1 axial images matched the EPI.

The motor and semantic generation tasks were analyzed using the BOLDfold technique, which involves correlating the raw data with the averaged BOLD function. This method of analysis requires that sufficient time elapse between tasks for the hemodynamic response function (HRF or BOLD function) to fully return to baseline levels. After correcting for baseline drift, the mean BOLD function for each voxel, collapsing across the repetitions of task and baseline, was empirically determined then repeated and correlated to the actual data as a measure of consistency across repetitions. In other words, the empirically determined BOLD function averaged over blocks was correlated to the actual data as a measure of consistency across repetitions. The squared correlation (r^2) represents the goodness of fit between the mean BOLD function and the observed BOLD data, capturing the variance accounted for in the data by the mean BOLD response. This method also serves to reduce the number of false activations associated with the traditional t test method, and, in particular, it is less sensitive to motion artifacts (Sarty and Borowsky, 2005). The correlation, r, was used as follows. A threshold correlation of $r = 0.60$ was used to define an active voxel. The false-positive probability is $p < 0.05$ with this threshold using a

Bonferroni-correction for 100,000 comparisons (the approximate number of voxels in an image volume). The use of both a gap in image acquisition and the BOLDfold analysis method minimizes motion artifact (see Sarty and Borowsky, 2005 for a detailed description).

FMRI maps were computed for the motor and semantic generation tasks using a technique for separating activations unique to each condition from those that are shared between conditions (Borowsky et al., 2005a, 2007). For each condition, C, for each participant, a threshold map $r_C(p)$ of r correlation values and a visibility map V_C(p) (intensity of BOLD amplitude) were computed where p is a voxel coordinate. The corresponding activation map for C, for each participant, was defined as $M_C(p) = \chi_{C,\theta}(p)V_C(p)$ where $\chi_{C,\theta}(p) = 1$ if $r_C(p) > \theta$ and zero otherwise. This threshold value represents the minimal acceptable correlation between the original BOLD function and its mean (repeated across the five blocks), and thus serves as a BOLD response consistency threshold. In other words, voxels are included into a binary mask if the correlation representing consistency between the original BOLD function and its mean exceeds 0.60 (binary value = 1), and excluded if the coefficient is below 0.60 (binary value = 0). We used a threshold of $\theta = 0.60$ to define active voxels. Shared maps (M_{shared}), and unique maps (M_{unique}) were computed for paired conditions A and B for each participant according to:

$$M_{unique}(p) = [\chi_{A,\theta}(p)V_A(p) - \chi_{B,\theta}(p)V_B(p)] \times [1 - \chi_{A,\theta}(p)\chi_{B,\theta}(p)] \quad (1)$$

$$M_{shared}(p) = \chi_{A,\theta}(p)\chi_{B,\theta}(p)[V_A(p) + V_B(p)]/2 \quad (2)$$

Equation 1 (unique activation) examines the two conditions for each voxel within each participant: if both conditions surpass the BOLD consistency threshold correlation, then the latter part of the equation amounts to [1–1] and the unique activation for that voxel would be zero; if only one condition surpasses the consistency threshold, then the latter part of the equation amounts to [1–0] and the unique activation for that voxel for that condition would be determined by the earlier part of the equation [BOLD intensity for condition A–0] or [0–BOLD intensity for condition B]. In other words, unique activation is driven by only one, but not both, conditions passing the BOLD consistency threshold. Equation 2 (shared activation) also examines the same two conditions for each voxel within each participant: if both conditions surpass the BOLD consistency threshold correlation, then the earlier part of the equation amounts to 1 * 1 and the shared activation for that voxel would be determined by the average of the BOLD intensities for both conditions; if only one condition surpasses the consistency threshold, then multiplication by 0 results in zero shared activation for that voxel. In other words, shared activation is driven by both conditions passing the BOLD consistency threshold.

The unique map represents a difference $(A \cup B) \setminus (A \cap B)$ and shows task subtraction for activations that are not common to conditions A and B (A is > 0, B is < 0). The shared map represents an intersection $A \cap B$ showing activation common to both conditions A and B with the activation amplitude coded as the average of A and B. Unique and shared maps were averaged across participants separately for each condition after smoothing and transformation to Talairach coordinates (Talairach and Tournoux, 1988) to produce the final maps as described below. Consistent and significant low-intensity BOLD functions are as important to understanding perception and cognition as consistent and significant high-intensity BOLD functions, thus the maps are presented without scaling the color to vary with intensity (i.e., the maps are binary, see also Borowsky et al., 2005a, 2006, 2007, 2012; Esopenko et al., 2008).

Using the AFNI software (Cox, 1996), voxels separated by 1.1 mm distance (i.e., the effective in-plane voxel resolution) were clustered, and clusters of volume less than 100 μL were clipped out at the participant level. The data were then spatially blurred using an isotropic Gaussian blur with a full width at half maximum (FWHM) of 3.91 mm. The averaging of images across subjects was subsequently done after Talairach transformation to a standardized brain atlas (Talairach and Tournoux, 1988). Visual inspection of the individual participant anatomical images did not reveal any structural abnormalities that would compromise the averaging of data in Talairach space. Mean activation maps in Talairach coordinates were determined for each map type along with the corresponding one sample t statistic for map amplitude (against zero) for each voxel. The final maps for the motor and semantic generation conditions surpass the θ threshold at an individual level, and the one-tailed t test of map amplitude against zero at the group level [$t(7) = 1.895$, $p < 0.05$].

RESULTS

Comparison of unique versus shared activation in the PFN is central to the hypotheses that are evaluated in this paper. A detailed description of the regions activated in all tasks can be found in **Figures 1**–**3** and listed in the Figure Captions. **Figures 1**–**3** clearly show both significant unique and significant shared activation in the comparison of the motor localization to pictures and words, for both hand and foot stimuli. A detailed description of all regions activated in the cortex for the motor localization and semantic generation tasks is also reported in **Figures 1**–**3**. The main finding of our experiment was that there is somatotopically organized shared activation in proximity to the sensorimotor and premotor cortices (see areas within the ellipses on **Figures 1B** and **D**, **2B** and **D**) for the processing of hand and foot semantic generation and motor localization tasks. Furthermore, somatotopically organized unique activation was shown for the hand and foot semantic generation task in proximity to the hand and foot motor localization tasks (see areas within the ellipses on **Figures 1A** and **C**, **2A** and **C**). In addition, in the comparison of the semantic generation of pictures versus words we found significant shared and unique activation in the ventral stream (**Figure 3**). This is to be expected given previous research showing that the ventral stream processes semantic information (for a review see Martin, 2001, 2007; Martin and Chao, 2001). For example, previous research has shown that generating action words to visually presented pictures and words activates the middle temporal gyrus (Martin and Chao, 2001), while the loss of conceptual object knowledge is associated with damage to the left posterior temporal cortex (Hart and Gordon, 1990).

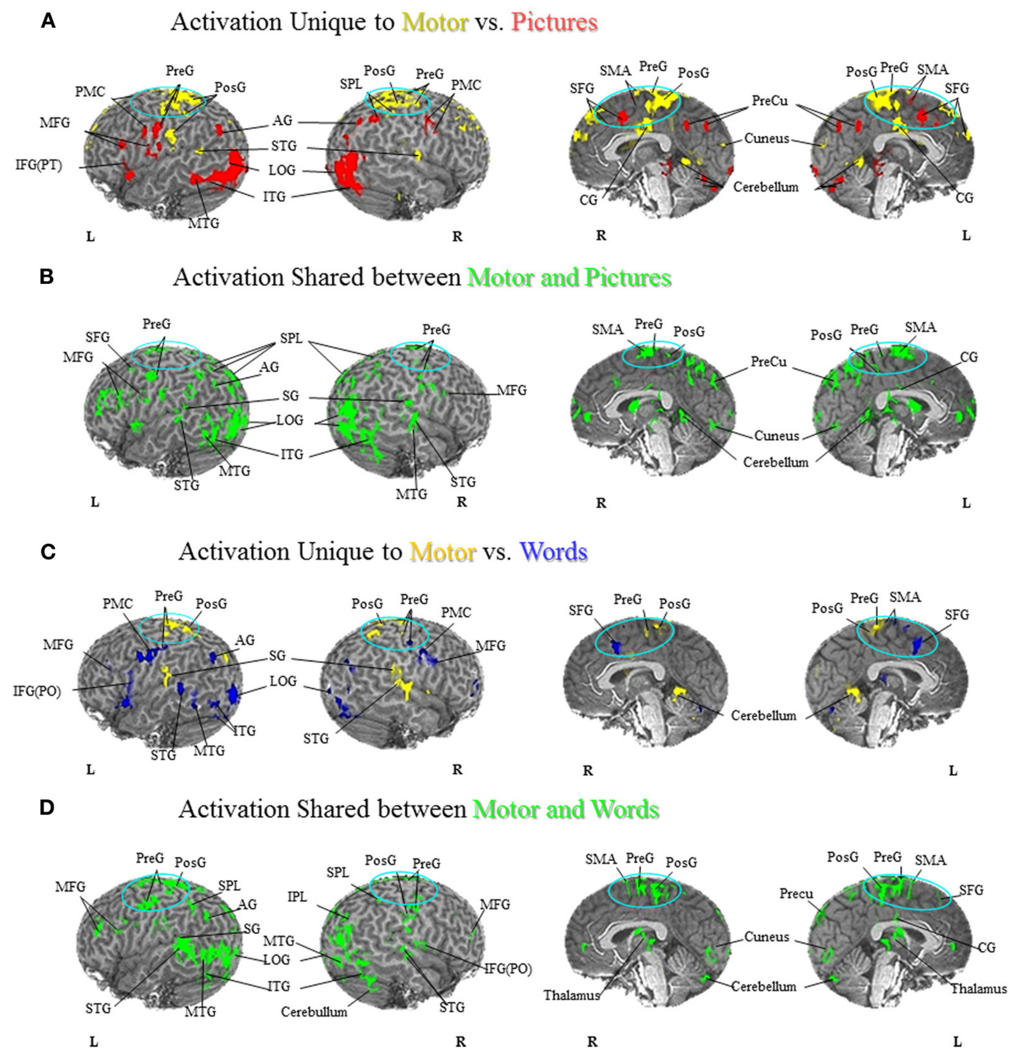


FIGURE 1 | Foot Motor Localization versus Semantic Generation. **(A)** Activation Unique to Motor vs. Pictures. Activation unique to the motor localization task (yellow coded activation) was found in bilateral superior temporal gyrus (STG), precentral gyrus (PreG) and postcentral gyrus (PosG), and the right superior parietal lobule (SPL) on the lateral surface. Activation was also found in bilateral PreG, PosG, supplementary motor area (SMA), superior frontal gyrus (SFG), and cingulate gyrus (CG) along the midline. For the semantic generation to foot picture stimuli (red coded activation), activation was found in bilateral premotor cortex (PMC) and angular gyrus (AG) on the lateral surface. Activation was also found in the left PreG, middle frontal gyrus (MFG), inferior frontal gyrus [pars triangularis; IFG(PT)], as well as the right SPL on the lateral surface. Activation along the midline was found in bilateral, SMA, SFG, and precuneus (PreCu). Regions of unique activation for the semantic generation to foot picture stimuli and motor localization task outside the PFN were found in bilateral inferior temporal gyrus (ITG) and lateral occipital gyrus (LOG), as well as the left middle temporal gyrus (MTG) on the lateral surface. Activation along the midline was found in bilateral cuneus and cerebellum. **(B)** Activation Shared between Motor and Pictures. Activation shared between the motor localization and semantic generation to foot picture stimuli (green coded activation) was found in bilateral PreG, SPL, and MFG, as well as the left SFG, supramarginal gyrus (SG), and AG on the lateral surface. Activation was also found in bilateral SMA, PreG, PosG, PreCu and CG along the midline. Regions of activation shared between the motor localization and the semantic generation to foot picture stimuli outside the PFN were found in bilateral ITG, MTG, STG, LOG, and right cerebellum on the lateral surface, and bilateral cuneus and cerebellum along the midline. **(C)** Activation Unique to Motor vs. Words. Activation unique to the motor localization task (yellow coded activation) was found in bilateral PreG, PosG, and SG, and right STG on the lateral surface. Activation was also found in bilateral PreG, and PosG, as well as the left SMA along the midline. For the semantic generation to foot word stimuli (blue coded activation), activation was found in bilateral PMC, MFG, and PreG, as well as left IFG [pars opercularis; IFG(PO)], and AG and right inferior parietal lobule (IPL) on the lateral surface. Activation was also found in bilateral SFG, and left SMA along the midline. Regions of unique activation for the semantic generation to foot word stimuli and motor localization task outside the PFN were found in the bilateral LOG and cerebellum, as well as the left STG, MTG, and ITG on the lateral surface. **(D)** Activation Shared between Motor and Words. Shared activation between the motor localization and semantic generation to foot word stimuli (green coded activation) was found in bilateral MFG, PreG, PosG, SPL, as well as the left AG and SG, and right IPL, and IFG(PO) on the lateral surface. Activation was also found in bilateral SMA, PreG, and PosG, and the left SFG, PreCu, and CG along the midline. Regions of shared activation between the motor localization and semantic generation to foot word stimuli outside the PFN were found in bilateral STG, ITG, MTG, LOG and cerebellum on the lateral surface. Regions of shared activation between the motor localization and semantic generation to foot word stimuli outside the PFN along the midline were found in bilateral cuneus, and cerebellum.

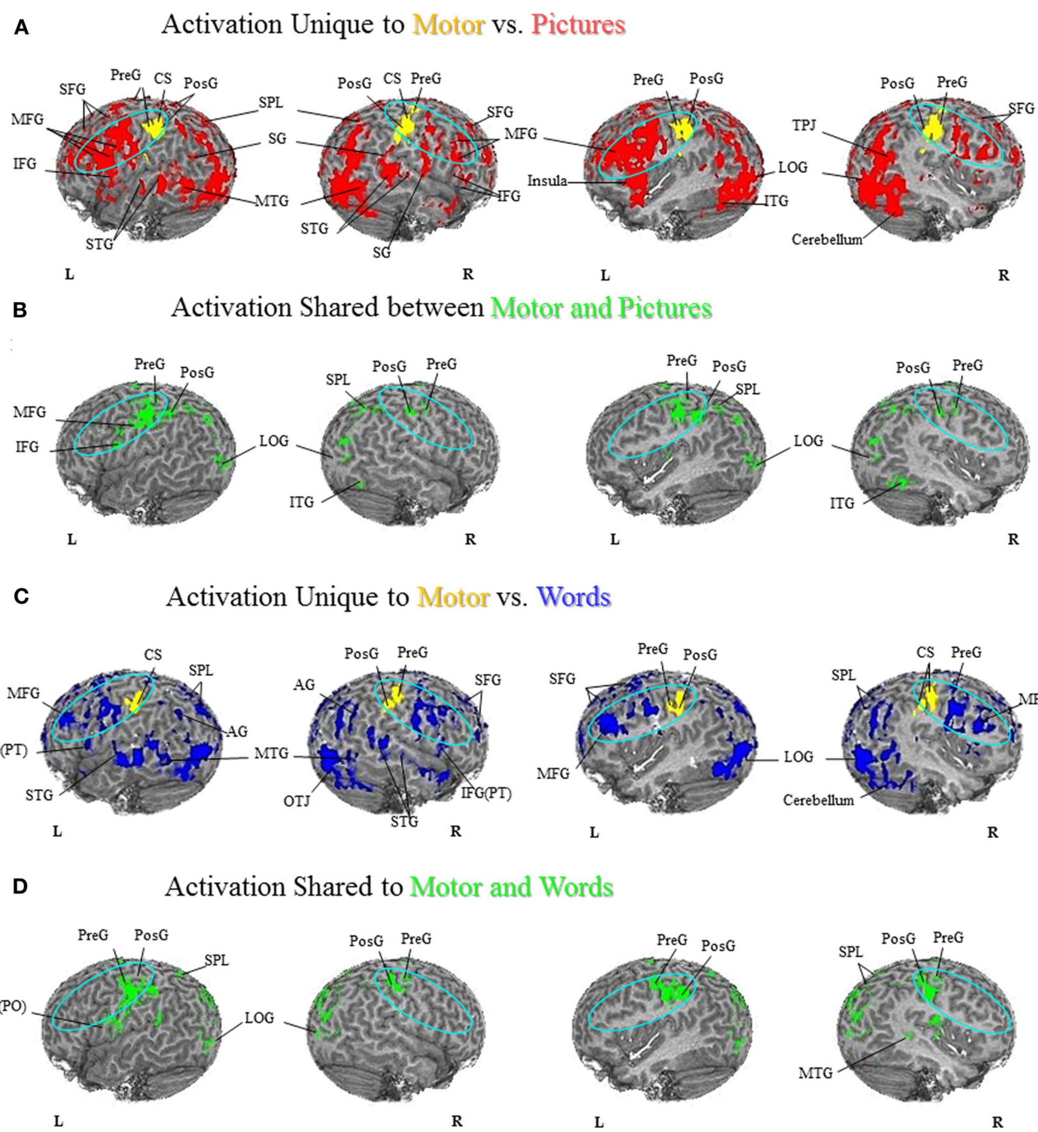


FIGURE 2 | Hand Motor Localization versus Semantic Generation. **(A)** Activation Unique to Motor vs. Pictures. Activation unique to the motor localization task (yellow coded activation) was found in bilateral PreG and PosG. For the semantic generation to hand picture stimuli (red coded activation), activation was found in bilateral MFG, SFG, PMC, IFG, SPL, and SG, as well as the right temporal-parietal junction, and IPL. Regions of unique activation outside the PFN for the semantic generation to hand picture stimuli were found in bilateral STG, MTG, and LOG, and right cerebellum. **(B)** Activation Shared between Motor and Pictures. Shared activation between the motor localization and semantic generation to hand picture stimuli (green coded activation) was found in bilateral PreG, PosG, and SPL, as well as the left PMC and IFG. Regions of shared activation between the motor localization and semantic generation to hand picture stimuli outside the PFN were found in the bilateral LOG and right ITG. **(C)** Activation Unique to Motor vs. Words. Activation unique to the motor localization task (yellow coded activation) was found in bilateral PreG and PosG. For the semantic generation to hand word stimuli (blue coded activation), activation was found in bilateral MFG, SFG, IFG(PT), SPL, and left AG, as well as right IPL. Regions of unique activation for the semantic generation to hand word stimuli outside the PFN were found in bilateral STG, MTG, LOG, as well as right cerebellum. **(D)** Activation Shared between Motor and Words. Shared activation between the motor localization and semantic generation to hand word stimuli (green coded activation) was found in bilateral PreG, PosG, and SPL, and the left SG and IFG(PO). Regions of shared activation between the motor localization task and semantic generation to words outside the PFN were found in bilateral LOG, and right MTG.

DISCUSSION

NEUROANATOMICAL EXAMINATION OF EMBODIED REPRESENTATIONS

The goal of our current research was to examine whether the semantic generation to pictures and words activated the PFN somatotopically, and moreover, to determine whether the semantic generation and motor localization tasks activated a shared network in the PFN. That is, to determine whether generating a use for a hand or foot stimulus activates regions in proximity to the sensorimotor cortices, and furthermore whether there

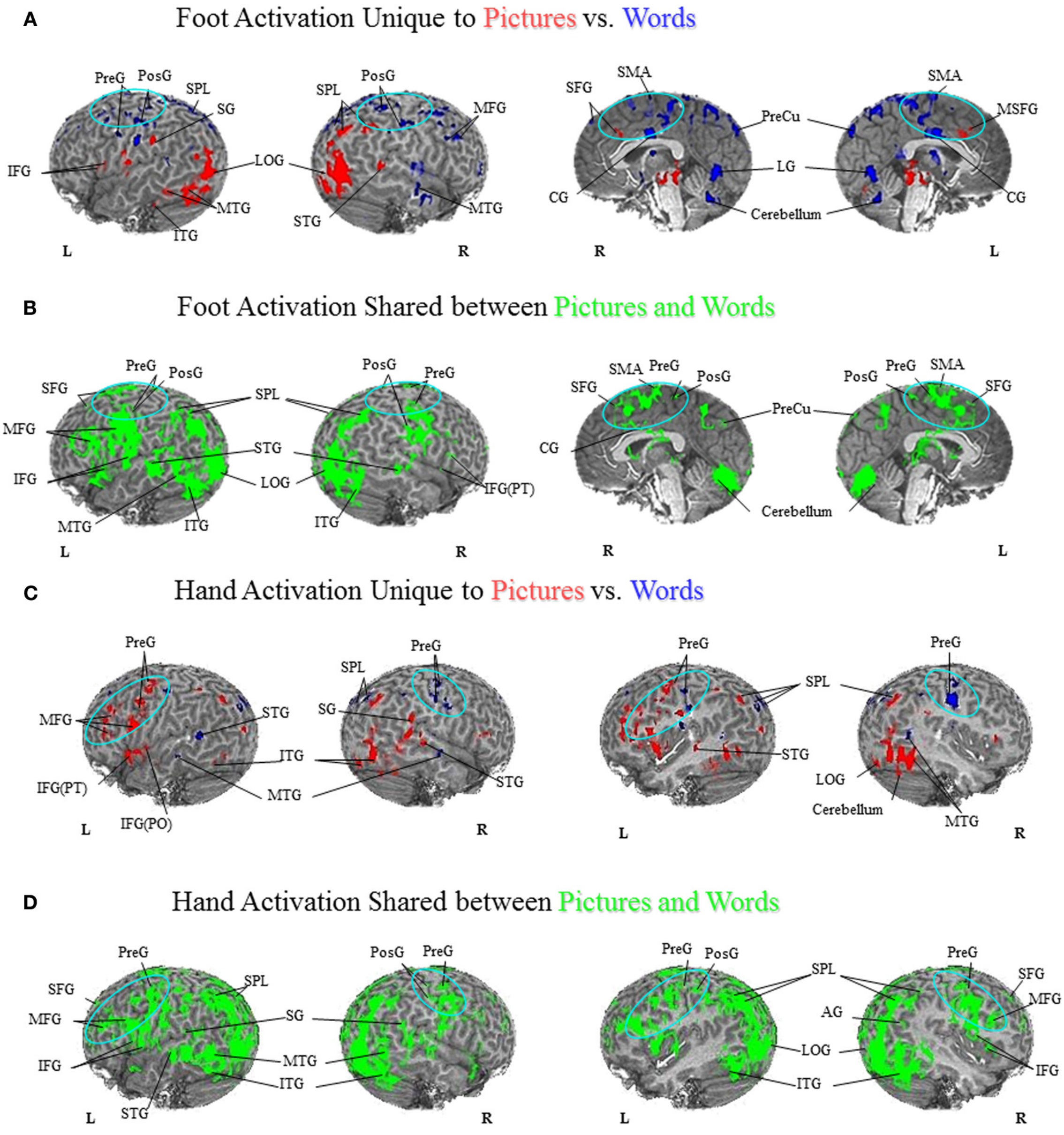


FIGURE 3 | Semantic Generation of Pictures versus Words. (A) Foot Activation Unique to Pictures vs. Words. For the semantic generation to foot picture stimuli (red coded activation), activation was found in the left IFG and SG and right IPL on the lateral surface, and bilateral SFG along the midline. For the semantic generation to foot word stimuli (blue coded activation), unique activation was found in bilateral PreG, PosG and SPL, and right MFG on the lateral surface, and bilateral SMA, CG, and PreCu, as well as right SFG along the midline. Regions of unique activation for the semantic generation to foot picture stimuli outside the PFN were found in the bilateral LOG, the left ITG and MTG, and the right STG on the lateral surface. Regions of unique activation for the semantic generation to foot word stimuli outside the PFN were found in the right MTG on the lateral surface, and bilateral lingual gyrus and cerebellum along the midline. **(B)** Foot Activation Shared between Pictures and Words. Shared activation between the semantic generation to foot picture stimuli and foot word stimuli (green coded activation) was found in bilateral PreG, PosG, PMC, IFG, and left SPL, AG, MFG, and SFG, on the lateral surface. Activation was also found in bilateral SFG, SMA, PreG, PosG, CG and PreCu, along the midline. Regions of shared activation for the semantic generation to foot picture and word stimuli outside the PFN were found in bilateral STG, ITG, and LOG, as well as the left MTG on the lateral surface, and bilateral cerebellum along the midline. **(C)** Hand Activation Unique to Pictures vs. Words. For the semantic generation to hand picture stimuli (red coded activation), activation was found in the left STG, MFG, PMC, IFG(PO), IFG(PT), PreG, the right SG, and bilateral SPL. For the semantic generation to hand word stimuli (blue coded activation), activation was found in bilateral PreG and SPL. Regions of unique activation for the semantic generation to hand picture stimuli outside the PFN were found in bilateral ITG, MTG, and the right STG, LOG, and cerebellum. Regions of unique activation for the semantic generation to hand word stimuli outside the PFN were found in the left STG, and bilateral MTG. **(D)** Hand Activation Shared between Pictures and Words. Shared activation between the semantic generation to hand picture stimuli and hand word stimuli (green coded activation) was found in bilateral MFG, IFG, SFG, PreG, PosG, PMC, SG, and SPL, as well as the left AG and right IPL. Regions of shared activation for the semantic generation to hand picture and hand word stimuli outside the PFN were found in bilateral STG, MTG, ITG, and LOG.

was shared activation between the semantic generation and motor localization tasks in proximity to the hand or foot sensorimotor cortices (see areas included in the ellipses in **Figures 1–3**). The word task in our experiment replicated earlier work from our lab showing unique and shared activation between the motor localization and semantic generation of words in the premotor regions. Our results show that, in the motor localization task, the foot-pedaling task produced unique activation in the motor cortex for feet (**Figures 1A,C**), while the finger-touch task produced unique activation in the motor cortex for hands (**Figures 2A,C**). In the comparison of the foot motor localization and semantic generation tasks, when participants responded to how they would interact with foot stimuli presented in both picture and word format there was a superior dorsal network of activation in the PFN (see areas in ellipses: **Figures 1A,C**). In the comparison of the hand motor localization and semantic generation tasks, when participants described how they would interact with hand stimuli presented in both picture and word format there was a dorsolateral network of activation in the PFN (see areas in ellipses: **Figures 2A,C**). As such, these results not only replicate earlier findings from the Esopenko et al. (2008) semantic generation task with words, but also show that generating responses to picture stimuli also activates regions proximal to the sensorimotor cortices. In addition, we found a shared network of activation between the hand and foot motor localization and semantic generation tasks in the PFN regardless of presentation format (i.e., pictures and words; see areas in ellipses: **Figures 1B,D** and **Figures 2B,D**). This shared activation was somatotopically organized in accordance with the sensorimotor somatotopic locations, whereby foot stimuli activated the dorsal regions of the PFN, while hand stimuli activated the dorsolateral regions of the PFN.

Embodied cognition theorists suggest that the brain represents semantic knowledge as a function of interacting with the environment, and to facilitate the processing of sensorimotor information (Wilson, 2002; Gibbs, 2006). According to Gallese and Lakoff (2005, pg. 456), an individual's "conceptual knowledge is embodied," whereby conceptual knowledge is "mapped within our sensory-motor system." Moreover, theories regarding mental simulation purport that conceptual processing is bodily based, in that it makes use of our sensorimotor system via simulation of action and perception (Svensson and Ziemke, 2004; Gallese and Lakoff, 2005). As such, the neural structures that are responsible for processing action and perceptual information would also be responsible for the conceptual processing of action-related language (Svensson and Ziemke, 2004; Grafton, 2009). Hence, one would expect that if the brain represents semantic knowledge in a way to facilitate sensorimotor processing, then we should see evidence of embodiment in the regions that process sensorimotor information (Barsalou, 1999). Past neuroimaging research has shown evidence consistent with embodied cognition, in that such research has demonstrated somatotopic semantic organization in the PFN when responding to action-related language (Hauk et al., 2004; Tettamanti et al., 2005; Boulenger et al., 2009). These results show that regions proximal to the motor cortex are activated when processing action-related language. However, to determine whether these regions reflect organization consistent with the theory of embodied cognition, we examined whether there is evidence of common regions of activation between motor and conceptual language tasks in the current study (see also Esopenko et al., 2008 experiment with word stimuli). The comparisons involving the hand and foot motor localization tasks and semantic generation to pictures and words showed that generating responses to hand and foot stimuli activated regions proximal to the sensorimotor and premotor cortices in a somatotopic fashion. Our results are consistent with the hypothesis that regions that encode sensorimotor experiences are activated when retrieving sensorimotor information, and thus provide neuroanatomical support for the theory of embodied cognition. In addition, these results show that the PFN responds to action-related stimuli regardless of word versus picture presentation format.

DISEMBODIED VERSUS EMBODIED THEORIES OF SEMANTIC MEMORY

A major criticism of embodied theories is that the motor system is activated as a by-product of the semantic analysis of a stimulus and is not required for semantic processing (Caramazza et al., 1990; Mahon and Caramazza, 2005, 2008). According to the disembodied view, conceptual representations are abstract and symbolic and are distinct entities from sensory and motor experiences (Caramazza et al., 1990; Mahon and Caramazza, 2005, 2008). However, in the disembodied view the motor system may still be activated, but it is not required. Specifically, although both conceptual and motor regions may be activated when processing conceptual information, processing occurs in conceptual regions and then spreads to the motor regions (Mahon and Caramazza, 2008). As such, conceptual processing is not associated with simulation of sensorimotor experiences, and moreover, is not a requirement to understand the meaning of a conceptual representation (Mahon and Caramazza, 2005, 2008). Evidence in favor of the disembodied account comes from apraxia patients who are impaired when using objects, but can name and pantomime the use associated with an object (Negri et al., 2007). This suggests that although there is damage to the motor regions, action-related language is still intact. However, it should be noted that the patients examined by Negri and colleagues have lesions that are not restricted to the motor, sensory, and parietal regions, but rather were wide-spread including regions outside the PFN (e.g., the temporal lobe). Nevertheless, disembodied theories suggest that activation in the premotor regions must be due to spreading activation from other regions after the semantic processing of a stimulus. That being said, even though the disembodied perspective can provide a plausible explanation as to why we see motor activation during conceptual processing, there is evidence that the sensorimotor and premotor cortices involvement during the processing of action-related semantic information is more than simply due to spreading activation.

Previous neuroimaging research has shown that the sensorimotor and premotor cortices are activated when processing action-related stimuli (e.g., silently reading action words or listening to action-related sentences), and that this activation is organized somatotopically dependent upon the effector the stimulus represents (Hauk et al., 2004; Pulvermuller, 2005; Tettamanti et al., 2005; Esopenko et al., 2008; Boulenger et al., 2009; Raposo et al., 2009; Boulenger and Nazir, 2010). In addition, behavioral research has shown that the sensorimotor properties of a sentence can affect an individual's ability to make a physical response to that sentence (Glenberg and Kaschak, 2002). Furthermore, behavioral studies have shown that the degree of physical interaction associated with a stimulus affects responding, with stimuli that are easier to interact with being responded to faster and more accurately in tasks that target semantic, phonological, and orthographic processing (Siakaluk et al., 2008a,b). Finally, studies of patient groups who have damage to the motor system and motor pathways show deficits in responding to action-related language, suggesting that the motor system is involved in responding to action-related semantic information (Bak et al., 2001; Boulenger et al., 2008; Cotelli et al., 2006a,b).

Recent electrophysiological and stimulation studies have provided some support for the theory of embodied cognition by demonstrating that the motor system is activated quickly, and likely during semantic processing and not just after it. To determine whether the sensorimotor and premotor regions are activated during or post semantic processing, previous research has used either: (1) magnetoencephalography (MEG) to examine whether semantic processing occurs before the sensorimotor system is activated, or whether the motor system is activated quickly following the presentation of a stimulus, which would suggest that semantic processing requires the motor system; or (2) by applying TMS to the motor system to determine whether responding to action words is facilitated or inhibited when stimulation is applied to these regions. Using MEG, Pulvermuller et al. (2005b) examined the spatial and temporal processing of spoken face-related (e.g., eat) and leg-related (e.g., kick) action words. They found that face-related and leg-related words activated the frontocentral and temporal regions. Of particular interest, Pulvermuller et al. found that the processing of face-related and leg-related words activated the frontocentral cortex somatotopically, whereby face-related words more strongly activated the inferior frontocentral regions, while leg-related words more strongly activated more dorsal superior central regions. Moreover, Pulvermuller and colleagues (2005b) found that semantic processing occurred early in these regions, in that the inferior frontocentral and superior central regions were found to be activated approximately 170–200 ms after presentation of word stimuli. They also found early activation peaking around 160 ms in the superior temporal regions, but suggest that this activation was likely related to phonological, acoustic and lexical processing rather than semantic processing. As such, the authors suggest that access to semantic information in the frontocentral motor regions occurs quite early, suggesting that activation in these regions is not likely occurring after semantic processing takes place. Furthermore, research has shown that semantic activation typically occurs later than 200 ms after stimulus onset. For example, Pulvermuller et al. (2000) have shown that differentiating between classes of verbs referring to different action types began 240 ms following the onset of an action word. Moreover, Pulvermuller et al. (1999) have shown that the semantic distinction between noun and verb word classes happens between 200–230 ms, again suggesting that the early activation in the motor cortices shown by Pulvermuller et al. (2005b) most likely occurs just prior to, or at least during, the semantic analysis of the stimulus.

Given the findings that the motor system is involved in the processing of language, Pulvermuller et al. (2005a) sought to examine whether applying stimulation (through TMS) to the motor system affects the processing of action-related language. Sub-threshold TMS was applied to hand and leg cortical areas while participants read arm-related (e.g., grasp) and leg-related (e.g., kick) words, pseudowords, and completed a lexical decision task. They found that applying TMS to motor regions facilitated responses to action words. In particular, the authors found that when TMS was applied to the arm motor regions, lexical decisions to arm stimuli were faster than lexical decisions to leg stimuli, whereas when TMS was applied to the leg motor regions, lexical decisions to leg stimuli were faster than lexical decisions to arm stimuli. Based on the finding that sub-threshold TMS facilitates responding to effector-specific action words, Pulvermuller and colleagues (2005a) proposed that the activation of the motor regions is not simply due to the motor regions being activated after semantic processing, but rather that these regions are actively involved in processing action-related language. Furthermore, they suggested that the sensorimotor regions process language information that is effector-specific, and thus play a significant role in the processing of effector-specific action words. Taken together, the findings from both studies suggest that the involvement of the motor system in the processing of action-related language is not simply a by-product of the semantic processing of the stimulus, but rather that the motor system plays a role in the semantic processing of the stimulus.

Our functional imaging results are consistent with the theory of embodied cognition, in that they show that the motor and premotor system is involved in responding to action-related stimuli. Specifically, the results demonstrated shared, or overlapping, activation in regions that are activated during a motor localization task and during a semantic generation task where no arm and leg motor movements occurred. The shared activation as measured here (within-participants and prior to averaging, unlike other conjoint analysis methods in the literature that do not first isolate the shared regions within-participants) between the motor and semantic tasks can be taken as support for embodied cognition in the spatial domain of brain topography, as it shows that activation of the motor system overlaps spatially with activation for conceptual representations. However, research still needs to be done to determine the temporal dynamics of this system using electrophysiological methods (e.g., event-related potentials, electroencephalography and MEG) during an overt semantic generation task. If the methods with higher temporal resolution ultimately demonstrate that semantic activation occurs prior to the activation in these shared regions (as measured by a motor task and an overt semantic generation task), then there would be more compelling evidence in support of disembodied cognition in the temporal domain of mental chronometry. That said, one must also use caution when employing this logic, as it ignores the issue of top-down processing. For example, activation that has been reported in primary visual cortex as a function of mental imagery (e.g., Kosslyn et al., 1995) can not begin before some degree of semantic activation has occurred (i.e., one needs to know what one is to imagine before the primary visual regions can simulate the referent), but it does not follow that this primary visual activation is just a by-product of imagination, when it clearly reflects top-down activation of an essential component of visual imagery. Later motor/somatosensory activation in the current context could simply reflect a top-down effect of semantics on the motor/sensory system, but need not make the involvement of the motor/sensory system any less interesting or important. Indeed, if one were to re-define embodied cognition as simply another example of imagery (or simulation), like visual imagery but in the motor/sensory domain, it would be part of a larger (and less-contentious) field of research on top-down processing.

PROCESSING OF PICTURE AND WORD STIMULI IN THE PFN.

Previous patient and behavioral research suggests that pictures and words have differential access to action-related knowledge (Lhermitte and Beauvois, 1973; Bub et al., 1988; Lambon Ralph and Howard, 2000; Chainey and Humphreys, 2002; Saffran et al., 2003; Thompson-Schill et al., 2006). Furthermore, previous neuroimaging research has shown that pictures and words are processed in both shared and unique brain regions (Borowsky et al., 2005a; Vandenberghe et al., 1996). Based on these results, we had predicted that if pictures and words both result in access to action-related semantic representations, we should see shared regions of activation in the PFN. However, given that patients can show a deficit in the ability to retrieve information when a stimulus is presented in picture format, but still have access to this same information when the stimulus is presented in word format (and vice versa), we predicted that we should also see unique activation in the PFN. As such, our research sought to examine whether picture and word stimulated action-related processes occur in the same regions, as shown by shared activation, or whether they are processed in separate regions, as shown by unique activation. As shown in **Figure 3**, there is substantial shared activation in the PFN between pictures and words, demonstrating that both stimuli formats have access to action representations in the PFN. However, our results also show unique activation between pictures and words in the PFN, suggesting that there is differential access to action-related knowledge. Taken together, our results suggest that there is both shared and unique activation for pictures and words in regions that process embodied information.

LIMITATIONS

One limitation of our research was that participants only responded to either arm or leg stimuli. Given that our analysis for computing unique and shared activation maps requires within-participant manipulation of conditions, we could not determine the degree to which (or if any) shared or unique activation seen in the PFN was due to overlap between effectors, or some general overlap in processing. In other words, it could be the case that responding to hand and foot stimuli could cause cross-effector activation, where some hand stimuli may activate foot regions, while some foot stimuli may activate hand regions. Such overlapping motor programs could potentially affect the shared and unique activation in the PFN. As such, one avenue for future research is to have each participant complete each of the hand and foot motor localization and semantic generation tasks. Such a design would allow us to compare all possible combinations of conditions, and determine whether any shared activation may be due to overlap between hand and foot motor programs.

CONCLUSION

The fMRI experiment presented here provides a comprehensive examination of a variant of a semantic generation task that permits participants to express their own semantic knowledge in response to action-related picture and word stimuli. Using this ecologically valid task, and a method of analysis that allows for a fair separation of shared regions of processing from unique regions, the functional neuroimaging results extend the data pertinent to evaluating the theory of embodied cognition. Sensorimotor and premotor regions are activated when openly responding to action-related stimuli, and there is shared activation between the motor localization tasks and the semantic generation tasks in the PFN, for both word and picture action-related stimuli.

ACKNOWLEDGMENTS

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada in the form of post-graduate scholarships to C. Esopenko and J. Cummine, a research assistantship to L. Gould, an undergraduate student research award to N. Kulhmann, and grants to G. E. Sarty and R. Borowsky. R. Borowsky is currently funded by NSERC Discovery Grant 183968–2008–2013. The lead author (C. Esopenko) and senior author (R. Borowsky) contributed equally to this work.

REFERENCES

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* 16, 1818–1823.

Bak, T. H., O'Donovan, D. G., Xuereb, J. H., Boniface, S., and Hodges, J. R. (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease-dementia-aphasia syndrome. *Brain* 124, 103–120.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behav. Res. Methods* 39, 445–459.

Barsalou, L. (2009). Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Biol. Sci.* 364, 1281–1289.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Borowsky, R., Cummine, J., Owen, W. J., Friesen, C. K., Shih, F., and Sarty, G. (2006). FMRI of ventral and dorsal processing streams in basic reading processes: insular sensitivity to phonology. *Brain Topogr.* 18, 233–239.

Borowsky, R., Esopenko, C., Cummine, J., and Sarty, G. E. (2007). Neural representations of visual words and objects: a functional MRI study on the modularity of reading and object processing. *Brain Topogr.* 20, 89–96.

Borowsky, R., Esopenko, C., Gould, L., Kuhlmann, N., Sarty, G., and Cummine, J. (2012). Localisation of function for noun and verb reading: converging evidence for shared processing from fMRI activation and reaction time. *Lang. Cogn. Processes Spec. Sect. Cogn. Neurosci. Lang.*

Borowsky, R., Loehr, J., Friesen, C. K., Kraushaar, G., Kingstone, A., and Sarty, G. (2005a). Modularity and intersection of "what", "where", and "how" processing of visual stimuli:

a new method of fMRI localization. *Brain Topogr.* 18, 67–75.

Borowsky, R., Owen, W. J., Wile, T. A., Friesen, C. K., Martin, J. L., and Sarty, G. E. (2005b). Neuroimaging of language processes: fMRI of silent and overt lexical processing and the promise of multiple process imaging in single brain studies. *Can. Assoc. Radiol. J.* 56, 204–213.

Boulenger, V., Hauk, O., and Pulvermuller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914.

Boulenger, V., Mechtouff, L., Thobois, S., Brousolle, E., Jeannerod, M., and Nazir, T. A. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia* 46, 743–756.

Boulenger, V., and Nazir, T. A. (2010). Interwoven functionality of the brain's action and language systems. *Ment. Lexicon* 5, 231–254.

Bub, D., Black, S., Hampson, E., and Kertesz, A. (1988). Semantic encoding of pictures and words: some neuropsychological observations. *Cogn. Neuropsychol.* 5, 27–66.

Caramazza, A., Hillis, A. E., Rapp, B. C., and Romani, C. (1990). The multiple semantics hypothesis: multiple confusions? *Cogn. Neuropsychol.* 7, 161–189.

Chainey, H., and Humphreys, G. W. (2002). Privileged access to action for objects relative to words. *Psychon. Bull. Rev.* 9, 348–355.

Cotelli, M., Borroni, B., Manenti, R., Alberici, A., Calabria, M., Agosti, C., Arevalo, A., Ginex, V., Ortelli, P., Binetti, G., Zanetti, O., Padovani, A., and Cappa, S. F. (2006a). Action and object naming in frontotemporal dementia, progressive supranuclear palsy, and corticobasal degeneration. *Neuropsychology* 20, 558–565.

Cotelli, M., Manenti, R., Cappa, S. F., Geroldi, C., Zanetti, O., Rossini, P. M., and Miniussi, C. (2006b). Effect of transcranial magnetic stimulation on action naming in patients with Alzheimer's disease. *Arch. Neurol.* 63, 1602–1604.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173.

Esopenko, C., Crossley, M., Haugrud, N., and Borowsky, R. (2011). Naming and semantic processing of action-related stimuli following right versus left hemispherectomy. *Epilepsy Behav.* 22, 261–271.

Esopenko, C., Borowsky, R., Cummine, J., and Sarty, G. (2008). Mapping the semantic homunculus: a functional and behavioral analysis of overt semantic generation. *Brain Topogr.* 22, 22–35.

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479.

Gibbs, R. W. (2006). *Embodiment and cognitive science*. New York, NY: Cambridge University Press.

Glenberg, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.* 9, 558–565.

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Ann. N.Y. Acad. Sci.* 1156, 97–117.

Hart, J., and Gordon, B. (1990). Delineation of single-word semantic comprehension deficits in aphasia, with anatomical correlation. *Ann. Neurol.* 27, 226–231.

Hauk, O. J. ohnsrude I., and Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Kosslyn, A. M., Thompson, W. L., Kim, I. J., and Alpert, N. M. (1995). Topographic representations of mental images in primary visual cortex. *Nature* 378, 496–498.

Lambon Ralph, M. A., and Howard, D. (2000). Gogi aphasia or semantic dementia? Simulating and accessing poor verbal comprehension in a case of progressive fluent aphasia. *Cogn. Neuropsychol.* 17, 437–465.

Lhermitte, F., and Beauvois, M. F. (1973). A visual-speech disconnexion syndrome: report of a case with optic aphasia, agnosia alexia, and colour agnosia. *Brain* 96, 695–714.

Mahon, B. Z., and Caramazza, A. (2005). The orchestration of the sensory-motor systems: clues from neuropsychology. *Cogn. Neuropsychol.* 22, 480–494.

Mahon, B. Z., and Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J. Physiol. Paris* 102, 59–70.

Martin, A. (2001). "Functional neuroimaging of semantic memory," in *Handbook of functional neuroimaging of cognition,* eds R. Cabeza and A. Kingstone (Cambridge, MA: MIT Press), 153–186.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.

Martin, A., and Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* 11, 194–201.

Negri, G., Rumiati, R., Zadini, A.,Ukmar, M., Mahon, B., and Caramazza, A. (2007). What is the role of motor simulation in action and object recognition? Evidence from apraxia. *Cogn. Neurosci.* 24, 795–816.

Penfield, W., and Rasmussen, T. (1950). *The Cerebral Cortex of Man: A Clinical Study of Localization of Function*. Oxford, England: Macmillan.

Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev.* 6, 576–81.

Pulvermuller, F., Harle, M., and Hummel, F. (2000). Neurophysiological distinction of verb categories. *Neuroreport* 11, 2789–2793.

Pulvermuller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005a). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797.

Pulvermuller, F., Lutzenberger, W., and Preissl, H. (1999). Nouns and verbs in the intact brain: evidence from even-related potentials and high-frequency cortical responses. *Cereb. Cortex* 9, 497–506.

Pulvermuller, F., Shtyrov, Y., and Ilmoniemi, R. J. (2005b). Brain signatures of meaning access in action word recognition. *J. Cogn. Neurosci.* 17, 884–892.

Raposo, A., Moss, H. E., Stamatakis, E. A., and Tyler, L. (2009). Modulation of motor and premotor cortices by actions, action words, and action sentences. *Neuropsychologia* 47, 388–396.

Saffran, E. M., Coslett, H. B., and Keener, M. T. (2003). Differences in word associations to pictures and words. *Neuropsychologia* 41, 1541–1546.

Sarty, G., and Borowsky, R. (2005). Functional MRI activation maps from empirically defined curve fitting. *Magn. Reson. Eng.* 24b, 46–55.

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., and Sears, C. R. (2008b). Evidence for the activation of sensorimotor information during visual word recognition: the body-object interaction effect. *Cognition* 106, 433–443.

Siakaluk, P. D., Pexman, P. M., Sears, C. R. Wilson, K., Locheed, K., and Owen, W. J. (2008a). The benefits of sensorimotor knowledge: body-object interaction facilitates semantic processing. *Cogn. Sci.* 32, 591–605.

Svensson, H., and Ziemke, T. (2004). "Making sense of embodiment: simulation theories and the sharing of neural circuitry between sensorimotor and cognitive processes,"in *Presented at the 26th Annual Cognitive Science Society Conference*, Chicago, IL.

Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York, NY: Thieme Medical Publishers, Inc.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., and Perani, D., (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.* 17, 273–81.

Thompson-Schill, S. L., Kan, I. P., and Oliver, R. T. (2006). Functional neuroimaging of semantic memory. in *Handbook of Functional Neuroimaging of Cognition 2nd edn*, eds R. Cabeza and A. Kingstone (Cambridge, MA: MIT Press), 149–190.

Vandenberghe, R., Price, C., Wise, R., Josephs, O., and Frackowiak, R. S. J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature* 383, 254–256.

Willems, R. M., and Hagoort, P. (2007). Neural evidence for the interplay between, language, gesture, and action: a review. *Brain Lang.* 101, 278–289.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 December 2011; accepted: 25 March 2012; published online: 207 April 2012.

*Citation: Esopenko C, Gould L, Cummine J, Sarty GE, Kuhlmann N and Borowsky R (2012) A neuroanatomical examination of embodied cognition: semantic generation to action-related stimuli. Front. Hum. Neurosci. **6**:84. doi: 10.3389/fnhum.2012.00084*

APPENDIX A

Stimuli (see Table 1, available on Frontiers of Neuroscience website: http://www.frontiersin.org/Human_Neuroscience/10.3389/fnhum.2012.00084/abstract)

APPENDIX B

List of Figure Abbreviations

AG	Angular gyrus
CG	Cingulate gyrus
IFG(PO)	Inferior frontal gyrus (pars operculatis)
IFG(PT)	Inferior frontal gyrus (pars triangularis)
IPL	Inferior parietal lobe
ITG	Inferior temporal gyrus
LOG	Lateral occipital gyrus
MFG	Middle frontal gyrus
MTG	Middle temporal gyrus
PMC	Premotor cortex
PreCu	Precuneus
PreG	Precentral gyrus
PosG	Postcentral gyrus
SFG	Superior frontal gyrus
SG	Supramarginal gyrus
SMA	Supplementary motor area
SPL	Superior parietal lobule
STG	Superior temporal gyrus

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 17 October 2012
doi: 10.3389/fnhum.2012.00285

A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery

Chris McNorgan*

Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, USA

Edited by:
Penny M. Pexman, University of Calgary, Canada

Reviewed by:
Ron Borowsky, University of Saskatchewan, Canada
Sharlene D. Newman, Indiana University, USA

****Correspondence:***
Chris McNorgan, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA.
e-mail: chris.mcnorgan@alumni.uwo.ca

The relationship between imagery and mental representations induced through perception has been the subject of philosophical discussion since antiquity and of vigorous scientific debate in the last century. The relatively recent advent of functional neuroimaging has allowed neuroscientists to look for brain-based evidence for or against the argument that perceptual processes underlie mental imagery. Recent investigations of imagery in many new domains and the parallel development of new meta-analytic techniques now afford us a clearer picture of the relationship between the neural processes underlying imagery and perception, and indeed between imagery and other cognitive processes. This meta-analysis surveyed 65 studies investigating modality-specific imagery in auditory, tactile, motor, gustatory, olfactory, and three visual sub-domains: form, color and motion. Activation likelihood estimate (ALE) analyses of activation foci reported within- and across sensorimotor modalities were conducted. The results indicate that modality-specific imagery activations generally overlap with—but are not confined to—corresponding somatosensory processing and motor execution areas, and suggest that there is a core network of brain regions recruited during imagery, regardless of task. These findings have important implications for investigations of imagery and theories of cognitive processes, such as perceptually-based representational systems.

Keywords: embodied cognition, imagination, imagery, modality-independent, modality-specific, semantic memory

Perception describes our immediate environment. Imagery, in contrast, affords us a description of past, future and hypothetical environments. Imagery and perception are thus two sides of the same coin: Perception relates to mental states induced by the transduction of energy external to the organism into neural representations, and imagery relates to internally-generated mental states driven by representations encoded in memory. Various forms of mental imagery have been implicated in a wide array of cognitive processes, from language comprehension (Bottini et al., 1994), to socially-motivated behaviors such as perspective taking (Ruby and Decety, 2001), to motor learning (Yágüez et al., 1998). Understanding the networks supporting imagery thus provides valuable insights into many behaviors.

WHAT ARE THE NEURAL SUBSTRATES OF MODALITY-SPECIFIC IMAGERY?

Though representations generated through mental imagery clearly have perceptual analogs, a persistent question of the imagery literature concerns the extent to which imagery and perceptual processes overlap. Within the visual imagery domain, Kosslyn and Thompson (2003) analyzed contemporary neuroimaging studies to explain the lack of consistency with which studies demonstrate recruitment of early visual cortex during imagery. They showed that imagery was most likely to recruit early visual cortex when it requires attention to high-resolution detail, suggesting that perceptual processing during imagery depends on attention or processing level (Craik and Lockhart, 1972). The analogs question has been posed in the auditory and motor imagery domains, with some studies finding activation in primary sensorimotor areas (Wheeler et al., 2000; Hanakawa, 2002; Bunzeck et al., 2005) and others not (Zatorre and Halpern, 1996; Halpern and Zatorre, 1999; Vingerhoets et al., 2002).

In an early review of the imagery literature, Kosslyn et al. (2001) concluded from the auditory and motor imagery that dominated literature at the time, that "most of the neural processes that underlie like-modality perception are also used in imagery," (p. 641). Subsequent study of imagery in other modalities and continuations of earlier lines of imagery study now afford a clearer picture of imagery across all sensory modalities and, importantly, of imagery in general. Moreover, recently developed analytic techniques now permit a more precise description of the perception-related processes underlying imagery. The present paper uses one such analytic technique to explore the body of modality-specific imagery literature with the overall aim of identifying the neural substrates of modality-specific and modality-general imagery. As will be discussed below, of particular importance is the question of whether modality-specific imagery recruits primary sensorimotor cortex as a rule. The resolution of this this question bears importantly on issues central to cognitive processes with which imagery is tightly bound.

IMAGERY AND PERCEPTUALLY-GROUNDED REPRESENTATIONS: THEORETICAL ISSUES

Semantic memory—one's knowledge of the meaning of things—critically supports a wide array of cognitive processes, from language production and comprehension, to action planning. Of all cognitive processes, imagery and semantic processing are perhaps most closely related. Imagery regularly relies on previously organized and stored semantic information (Kosslyn et al., 1997) about the features to be imagined. A large body of literature makes the complimentary argument that the reactivation of perceptual representations—that is, imagery—underlies semantic retrieval. The assumption that imagery underlies semantic retrieval is the central premise of perceptually-based theories of cognition. The Perceptual Symbol System account (Barsalou, 1999) assumes that reactivation of perceptual representations ("perceptual simulations") underlies semantic retrieval and provides one of the most recent and explicit accounts of the importance of imagery to semantic processing. Under this account, perceiving an object elicits a unique pattern of activation in primary sensorimotor cortices encoding salient perceptual properties of that object. Perceptually-based theories argue that encoding and retrieving these activations within the perceptual system naturally permits high-fidelity perceptually-rich representations. Similar ideas underlie Warrington and McCarthy's sensory/functional theory (Warrington and McCarthy, 1987), and Paivio's dual coding theory (Paivio, 1971), which explicitly argues that abstract propositional and (visual) imagery representations comprise concept knowledge.

Full elucidation of the assumptions and criticisms of a perceptually-grounded system are beyond the scope of this article, but have been given extensive consideration elsewhere (Barsalou, 1999; Simmons and Barsalou, 2003). One advantage of perceptually-grounded models is that they arguably overcome the reverse inference problem (Poldrack, 2006), which is the neuroimaging equivalent of the symbol grounding problem (Harnad, 1990). The symbol grounding problem describes the circularity inherent in relating arbitrary symbols to an equally arbitrary symbol system. The solution Harnad proposes is for one symbolic system to be non-arbitrary—that is, to be grounded in an external physical system. Because primary sensory cortices contain populations driven by external physical systems, the perceptual system provides the grounding required to understand those cognitive systems that interact with it. For example, patterns of activations within olfactory cortex reflect detection of particular smells. Olfactory imagery and knowledge retrieval might engage a wide network of brain areas related to any number of cognitive processes. However, if imagery is perceptually-grounded, one would additionally expect an involvement of the corresponding sensory cortex. Whatever other brain regions may contribute toward olfactory imagery, it is relatively straightforward to argue activity within olfactory cortex is part of an olfactory representation.

IMAGERY AND PERCEPTUALLY-GROUNDED REPRESENTATIONS: METHODOLOGICAL ISSUES

The strong theoretical ties between perceptually-based semantic theories and imagery suggest that a thorough understanding of the former requires an understanding of imagery. It is important to reiterate that perceptually-based representational theories assume that semantic representations are rooted in imagery, rather than perception *per se*. Nonetheless, a common practice is to localize these perceptually-based representational systems using perceptual tasks. For example, Simmons et al. (2007) investigated color knowledge retrieval within color-sensitive visual cortex localized using a modified Farnsworth-Munsell 100-Hue Task (Farnsworth, 1957).

The demonstration of a common neural basis underlying perception and modality-specific semantic knowledge provides compelling support for perceptually-based theories. Such findings support a strong version of a perceptually-grounded semantic system—that is, that perceptual processing is implied by semantic retrieval. The approach of using primary sensory cortex to define these representational areas has some limitations, however. First, the choice of localizer task is not a trivial consideration, and may impact the ability to detect the true extent of the modality-specific region. For example, recruitment of color-selective areas has been shown to be modulated by attention (Beauchamp et al., 1999), and thus different perceptual localizer tasks may give different estimates of the corresponding perceptual areas. Second, multiple localization tasks and specialized delivery apparatus required for some perceptual tasks may be impractical for investigations of multiple representational modalities. Even when primary sensory regions are well-defined, there remains one important consideration: Semantic encoding and retrieval processes are assumed to be rooted in imagery. Thus, to the extent that the network supporting imagery extends beyond primary somatosensory perceptual areas, important imagery-related contributions to semantic encoding and retrieval may be overlooked. Thus, an understanding of the neural substrates underlying imagery provides critical insight into the organization of the semantic system, and can guide investigations of representational systems.

THE ALE META-ANALYTIC TECHNIQUE

An empirically-driven characterization of the neural correlates of modality-specific and modality-general imagery processes has been made possible in recent years by the development of meta-analytic techniques for assessing neuroimaging data. Techniques such as Activation Likelihood Estimation (ALE) (Chein et al., 2002; Turkeltaub et al., 2002) and Multilevel Kernel Density Analysis (MKDA) (Wager et al., 2007) allow the application of statistical measures to the literature to assess the reliability with which an effect is demonstrated in a particular brain area. In short, these methods permit an empirical test of consensus within a body of neuroimaging literature. A detailed explanation of the advantages and underlying statistics behind voxel-based meta-analytic approaches was presented by Laird et al. (2005). Briefly, these approaches examine the activation foci reported for a common contrast among multiple studies. Statistical tests on these data (e.g., chi-square analyses, Monte Carlo simulations) provide quantifiable, statistically-thresholded measures of the reliability of activation for a given contrast within a given region. As with other meta-analytic techniques, these approaches importantly highlight commonalities among studies, and minimize idiosyncratic effects. The ALE approach has been used in recent years to examine representational knowledge in the semantic system in general (Binder et al., 2009) and for more specific representational knowledge about categories such as tools and animals (Chouinard and Goodale, 2010). The utility of this approach in identifying important networks within these domains suggests it may be similarly useful in the conceptually-related imagery domain.

What follows is an ALE analysis of the neuroimaging literatures in modality-specific imagery across visual, auditory, motor, tactile, olfactory, and gustatory modalities. These analyses provide a descriptive survey of the imagery literature and were intended to meet three main goals: First, to identify the brain areas recruited during imagery, regardless of modality. Second, to identify within each modality the brain regions associated with modality-specific imagery with particular attention to the extent to which primary sensorimotor perceptual regions are recruited. Finally, various sub-processes are carried out by different and well-defined populations of neurons tuned for processing color, form and motion during visual perception. The number of studies investigating corresponding subtypes of visual imagery provides an opportunity to investigate whether evidence for a similar organization can be found during visual imagery.

MATERIALS AND METHODS

Searches for candidate imagery studies were conducted in the PubMed and Google Scholar databases for fMRI and PET studies related to imagery or sensory-specific imagery (e.g., “gustatory imagery,” “taste imagery”). Iterative searches within the citations among candidate imagery studies located additional candidate imagery studies with the intention of creating a comprehensive list of studies explicitly examining imagery or imagery-like tasks. For purposes of this study, imagery-like tasks were defined as those for which the retrieval of perceptual information from long term memory was required. These tasks were framed as perceptual knowledge retrieval by study authors and extensively cited or were cited by explicit studies of imagery. As discussed above, perception-based theories of knowledge representations are explicitly rooted in imagery (Paivio, 1971; Warrington and McCarthy, 1987; Barsalou, 1999), and a large body of literature supports the hypothesis that imagery underlies perceptual knowledge retrieval. Consequently, in many cases, similar tasks were used by different authors to investigate perceptual knowledge retrieval and imagery [e.g., color feature verification used by Kellenbach et al. (2001) and color feature comparison used by Howard et al. (1998)]. ALE measures concordance among reported activations; therefore heterogeneity among studies should lead to a reduction in power, rather than inflation of type I error. To the extent that perceptual knowledge retrieval does not involve imagery processes, inclusion of perceptually-based knowledge studies should therefore lead to slightly more conservative estimates of imagery activation. These studies comprised a small minority of the overall body of literature surveyed, however, so any such conservative bias should be rather small. For these reasons, these inclusionary criteria were deemed appropriate. Studies investigating special populations (e.g., synaesthetes, neurological patients) were excluded, as were those that did not conduct whole brain analyses or report coordinates in stereotactic space for significant modality vs. baseline contrasts. The studies included in the present analysis are listed in **Table 1**.

Imagery vs. low-level baseline contrasts were categorized with respect to one of eight modality conditions: Auditory, Tactile, Motor, Olfactory, Gustatory, Visual-Form, Visual-Color, and Visual-Motion. Modality categorizations were generally straightforward to determine (e.g., taste recall clearly relating to gustatory imagery), though classification of visual imagery subtypes required careful consideration of the task, stimuli and baseline contrasts used. One study (Roland and Gulyás, 1995) required participants to recall both the colors and geometric description of colored geometric patterns. The remaining visual form studies used monochromatic stimuli. The relative scarcity of color imagery studies, and the saliency of both form and color information in the task motivated the inclusion of this study in both modalities. Despite this single commonality, the ALE maps for these two modalities did not resemble one another.

The focus on the lowest-level baseline contrast was mandated by the fact that it alone was included across all imagery studies. Though the baseline task varied among studies, ranging from rest baselines to passive viewing baselines controlling for other modalities (e.g., passive viewing of scrambled scenes for auditory imagery) or within-modality (e.g., passive viewing of letter strings for form imagery), no particular baseline task dominated any modality. Implicit or resting-baselines were used in approximately 40% of studies, nearly all of which employed tasks requiring no overt response on the part of the participant. The remaining studies employed somewhat more complex baseline tasks generally designed to account for attention or response processes (e.g., those associated with button presses) under the assumption of cognitive subtraction. Direct contrasts between perception- and imagery-related activity tend to show reduced activity in primary perceptual areas for imagery relative to perception (e.g., Ganis, 2004). Care was thus taken to ensure that baselines involving a sensorimotor component excluded activity only in modalities of non-interest. For example, the detection of taste within a tasteless solution (Veldhuizen et al., 2007) likely involves motor activity in the planning and execution of passing the solution over the tongue and swallowing. The baseline task used in that study involved swallowing the solution without making a taste judgment. Under the assumption of pure insertion, the contrast between the two tasks should reveal activations associated only with the gustatory judgment (but see Friston et al., 1996 for a critique of the logic of cognitive subtraction). In the analyses that follow, however, the lack of systematicity among active baseline tasks somewhat mitigates concerns about the validity of cognitive subtraction. Aggregated across studies, imagery-related activations should be more reliable than those related to particular baseline choices, just as random-effects analyses across participants distinguish the influence of an experimental manipulation from noise. Though the complexity of baseline task was generally commensurate with that of the experimental task across studies, baseline complexity was investigated in detail in the general imagery analysis, where the numbers of studies permitted such an analysis. The reduced number of studies available for

Table 1 | Studies used in the imagery meta-analysis.

Modality	First author	Year	Imagery task	Baseline
AUD*	Belardinelli	2009	Imagine performing sensory action	Listen to sentence designating abstract concept
AUD	Bunzeck	2005	Imagined sounds corresponding to movie	Passive viewing of scrambled scene
AUD	Halpern	1999	Imagine continuation of tone sequence	Passive listening to tones
AUD*	Kellenbach	2001	Retrieval of sensory specific object knowledge concerning color, sound, size	Visual search for X in unrelated letter string
AUD*	Kiefer	2008	Lexical decision on stimuli with and without auditory features	Implicit/rest
AUD	Nyberg	2000	Recall sounds paired with textual cue	Implicit/rest
AUD	Wheeler	2000	Recall studied complex picture or sound	Opposite modality recall
AUD	Yoo	2001	Imagine recorded chord in response to cue	Implicit/rest
GUS*	Belardinelli	2009	Imagine performing sensory action	Listen to sentence designating abstract concept
GUS	Kikuchi	2005	Imagine taste of strong-tasting pictured foods	Viewing colored balls
GUS	Kobayashi	2004	Taste recall for pictured food items	Implicit/rest
GUS	Small	2003	Same/different judgments of pictured foods vs. locations	Passive viewing of locations
GUS	Veldhuizen	2007	Detection of taste in a tasteless solution	Passive swallowing
MTR*	Belardinelli	2009	Imagine performing sensory action	Listen to sentence designating abstract concept
MTR	Canessa	2007	Judgments whether items are manipulated using the same action	Implicit/rest
MTR	Creem-Regehr	2007	Mental rotation of self(motor) or other (visual)	Implicit/rest + no rotation
MTR	Dechent	2004	Imagined execution of trained finger tapping sequence	Visual imagery of scene
MTR	Guillot	2009	Imagined motor execution	Passive tone listening
MTR	Hanakawa	2002	Imagined execution of trained finger tapping sequence	Fixation
MTR*	Hauk	2004	Reading action words associated with specific body parts (e.g., "KICK")	Fixation
MTR	Johnson	2002	Imagined grip	Foil trials
MTR	Nyberg	2001	Imagined execution of actions	Implicit/rest
MTR	Servos	2002	Imagined execution of motor sequence	Visual object imagery
MTR*	Simmons	2003	Property verification	Lexical decision
MTR	Vingerhoets	2002	Mental rotation of tools and hands to make same/different judgments	Passive viewing (non-rotated pictures)
OLF*	Belardinelli	2009	Imagine performing sensory action	Listen to sentence designating abstract concept
OLF	Djordjevic	2005	Imagined odors	Odor detection in the absence of odor
OLF	Gottfried	2004	Recall of odor paired with object pictures during training	View picture without associated odor
OLF	Plailly	2012	Odor imagery	Implicit/rest
OLF	Yeshurun	2009	Recall remembered smell	Implicit/rest
TAC*	Belardinelli	2009	Imagine performing sensory action	Listen to sentence designating abstract concept
TAC*	Newman	2005	Haptic/form judgments on pairs of concrete object names	Implicit/rest
TAC	Yoo	2003	Imagined tactile stimulation	Implicit/rest
VCO	Hsu	2011	Color word similarity judgment	Implicit/rest
VCO	Hsu	2012	Relative luminance decision on chromatic/achromatic object names	Evaluative judgments on abstract concepts
VCO*	Kellenbach	2001	Retrieval of sensory specific object knowledge	Visual search for X in unrelated letter string
VCO	Sack	2002	Mental clock task; color and angle judgments	Implicit/rest
VFO*	D'Esposito	1997	Visualize named concrete objects	Abstract concept
VFO	Ganis	2004	Visualize a line drawing	Implicit/rest
VFO	Gulyás	2001	Visualize capital letters from a known passage of text	Implicit/rest
VFO	Ishai	2000	Visualize recently studied or famous faces	Passively view letter strings
VFO*	Kellenbach	2001	Retrieval of sensory specific object knowledge	Visual search for X in letter string
VFO	Kosslyn	1993	Visualize uppercase block letters	Response to unrelated target
VFO	Kosslyn	1995	Visual form judgments on imagined line drawings	Passive listening
VFO	Kosslyn	1997	Visualize uppercase block letters	Response to target grid element
VFO*	Mellet	1996	Three-dimensional object visualization	Passive listening/rest

(Continued)

Table 1 | Continued.

Modality	First author	Year	Imagery task	Baseline
VFO*	Newman	2005	Haptic/form judgments on pairs of concrete object names	Implicit/rest
VFO*	Oliver	2009	Property verification	Lexical decision
VFO	Sack	2002	Mental clock task; color and angle judgments	Implicit/rest
VFO	Thompson	2001	Compare visualized and displayed patterns	Response to unrelated auditory cue
VFO	Trojano	2000	Comparing visualized clock faces	Numerical judgment
VFO	Yomogida	2004	Object imagery and synthesis	Implicit/rest
VMO	Alivisatos	1997	Mental rotation	Discrimination of retters/rumbers
VMO	Barnes	2000	Mental rotation and linear translation	Implicit/rest
VMO	Creem-Regehr	2007	Mental rotation of self (MTR) or other (VMO)	Implicit/rest + no rotation
VMO	de Lange	2005	Judge laterality of left/right hands	Visual imagery
VMO	Goebel	1998	Imagine previously studied moving stimuli	Implicit/rest
VMO	Guillot	2009	Visualized motor execution from 1st person perspective	Passive tone listening
VMO	Jordan	2001	Mental rotation	Same/different/ numerosity judgments on static figures
VMO	Kaas	2010	Imagine moving ball	Passively listening to unrelated tone
VMO	Slotnick	2005	Mental rotation	Passively attend to display

Note: "" denotes experiment with a semantic component; AUD, auditory, GUS, gustatory; MTR, motor; OLF, olfactory; TAC, tactile; VCO, visual-color; VFO, visual-form; VMO, visual-motion.*

individual modalities, however, precluded such an analysis within each modality.

Concordance among imagery vs. baseline activation foci reported across the neuroimaging literature was analyzed using a widely used activation likelihood estimate (ALE) meta-analytic approach (Eickhoff et al., 2012). Analyses were performed using GingerALE 2.1 (http://brainmap.org/ale/). Correction for multiple comparisons was performed using a false-discovery rate (FDR) threshold of $pN < 0.05$. GingerALE reports the number of voxels meeting the selected FDR threshold within each ALE map. Except where noted, a cluster size threshold, equal to the FDR rate times the number of suprathreshold voxels, was applied to each map (hereafter *extent-thresholded clusters*). For example, if 1000 voxels reached a FDR threshold of 0.05, then the expected number of false positives within that ALE map would be 50. A cluster size threshold of 50 in this example ensures that no extent-thresholded cluster would consist entirely of false positives. Because the number of FDR-significant voxels varied by modality, this approach resulted in different cluster thresholds across modalities. It should be noted, however, that imagery-related clusters were analyzed independently and with respect to sensorimotor ROIs rather than with each other (see below). Thus, these differences had little to bear on the results that follow, other than to increase the confidence with which conclusions can be drawn about the meaningfulness of any given extent-thresholded cluster for that analysis.

ROI DEFINITION AND OVERLAP ANALYSIS

The question of whether modality-specific imagery activates primary sensorimotor cortex was addressed within each modality by assessing the overlap between extent-thresholded ALE clusters and the primary sensorimotor ROI defined for each modality. ROIs were drawn from several publicly available anatomical atlases. The source(s) for each ROI are indicated in each modality analysis. Multiple atlases were necessitated by the fact that no single atlas contained ROI definitions corresponding to all modalities included in the present analysis. In some cases, different atlases contained different definitions of the same region. When a given anatomical region was defined in exactly one atlas, that definition was taken as the ROI; when multiple atlases defined the same region, the intersection (i.e., only those voxels common to all definitions) was taken as the ROI. This atlas-based approach was intended to arrive at a set of ROIs that are easily reproducible and for which there should be general agreement are representative of the corresponding sensorimotor cortices.

The degree of overlap was assessed for each ROI by determining whether the number of voxels in the extent-thresholded ALE clusters overlapping with a given ROI reached an overlap criterion. The overlap criterion was set independently for each ROI using 3dClustSim (available as part of the AFNI fMRI analysis package, available at http://afni.nimh.nih.gov/afni/download). Briefly, 3dClustSim calculates cluster size threshold (k) for false positive (noise-only) clusters at specified uncorrected alpha level. Though the ALE analyses used FDR corrected alpha thresholds, the equivalent voxel-wise alpha threshold for each ALE map is available in the GingerALE output. 3dClustSim carries out a user-specified number of Monte Carlo simulations of random noise activations at a particular voxel-wise alpha level within a masked brain volume. Ten thousand such simulations for each ALE map were used for this study. The number of simulations in which clusters of various sizes appear within the volumetric mask is tallied among these simulations. These data are then used to calculate size thresholds across a range of probability values for that region. For example, in a specified volume using a voxel-wise alpha of 0.001, if clusters of size 32 mm^3 or greater appear in 50 of 10,000 iterations by chance, this correspond to a $p < 0.05$ cluster-level significance threshold. In other words, within the specified volume using a voxel-wise alpha of 0.001, clusters exceeding $32\,mm^3$ are unlikely to occur by chance. To be clear, the cluster thresholds calculated using 3dClustSim was used to calculate an overlap criterion for each ROI, and not as an additional ALE cluster thresholding step. To the author's knowledge, no previous meta-analysis of neuroimaging data has attempted to qualify overlap between ALE clusters and a priori ROIs. However, the cluster size threshold approach is widely used to test statistical significance of clusters in conventional ROI analyses. That is, size thresholding is often used to determine whether a cluster of a particular size occurring within a given ROI is statistically significant. The present analysis had identical requirements, thus it was deemed to be an appropriate metric of overlap significance. A benefit of this approach when considering different ROIs is that it naturally takes into account differences in ROI extents: Larger sensorimotor ROIs require correspondingly greater overlap with imagery clusters for the overlaps to reach statistical significance.

Finally, it is important to note that the following analyses identify concordance of activation across studies within each modality, rather than contrast modalities directly. That is, they do not identify regions of activation unique to imagery in a particular modality. There are regions for which only studies of imagery for one modality converges (e.g., gustatory cortex activation apparent only for gustatory imagery studies). Nonetheless, the following results do not speak to whether one imagery modality recruits a particular region more than any other imagery. Inter-modal contrasts were not performed for two reasons: First, such contrasts address the question, not of what regions are implicated in a particular type of imagery, but what regions are implicated more for that type of imagery than any other. Networks defined by such contrasts would thus be more exclusive, and reducing the usefulness of these analyses to those interested in a non-comparative description of imagery for a particular modality. Second, there is a practical problem imposed by the disparity between the frequencies with which imagery in each sensorimotor modality has been investigated. This disparity would plausibly skew any such comparisons and generate networks driven by a single modality. When analyses are restricted to within-modality, however, differences with respect to numbers of studies are not problematic: a coherent network can be identified from relatively few studies, provided they are mutually consistent. Though modality-specific activations are not explicitly contrasted, crossmodal overlap between clusters is noted where it occurs.

RESULTS

The results are presented in order of generality. The first analysis identifies those regions consistently active relative to baseline in neuroimaging studies of imagery across all modalities. The eight subsequent analyses identify regions consistently active relative to baseline in modality-specific imagery for each of 5 sensorimotor modalities and 3 subtypes of visual imagery. All coordinates are reported in MNI standard space.

GENERAL IMAGERY NETWORK

A statistical threshold of $pN < 0.01$ (FDR corrected) and a minimum cluster size threshold of $800\,mm^3$ was used for the general imagery analysis. One thousand hundred and three foci from 84 contrasts involving 915 participants contributed toward these results. Nine primarily left-lateralized clusters reached the significance threshold (**Table 2**, **Figure 1**). These activations were found in bilateral dorsal parietal, left inferior frontal and anterior insula regions.

As indicated earlier, one advantage of meta-analytic techniques is that random-effects analyses minimize spurious effects attributable to idiosyncratic experimental design decisions among studies (e.g., choice of baseline) and highlight commonalities among them (e.g., choice of imagery modality). Imagery vs. baseline contrasts in the ALE analyses involved two broad classes of low-level baseline tasks: The resting state baseline tasks are assumed to be homogeneous across the 33 contrasts that used them. The non-resting state baseline tasks used across the remaining 50 contrasts were more varied, typically involving passive perceptual control conditions (for non-target modalities) or foil trials. Because ALE is sensitive to activation consistencies, it was plausible that baseline-related (rather than strictly imagery-related) networks may emerge in the ALE statistics. This concern was conservatively addressed by a conjunction analysis of resting-baseline vs. non-resting-baseline studies. The significance threshold was maintained at $pN < 0.01$ (FDR corrected) for both baseline conditions, though no cluster extent threshold was used (the resulting false discovery rate was 0.0001). The conjunction analysis found nine clusters, primarily in bilateral dorsal parietal and left inferior frontal regions that were active across all imagery modalities for all baseline conditions (**Table 2**). These results are suggestive of a core imagery network, though the extent of activation beyond this core network presumably depends baseline task. As the remaining analyses indicate, these activations also depend on imagery modality.

AUDITORY IMAGERY

Minimum cluster size threshold in the ALE analysis of auditory imagery studies was set at $632\,mm^3$. Ninety-three foci from 11 experiments involving 127 participants contributed toward these results. For the purposes of this analysis, primary auditory cortex was defined by the AAL template definition of Heschl's Gyrus within the MRIcroN software package (http://www.mccauslandcenter.sc.edu/mricro/mricron/index.html). Ten clusters were reliably associated with auditory imagery at a statistical threshold of $pN < 0.05$ (FDR corrected) (**Table 3**). No cluster overlapped with primary auditory cortex. Seven of eleven auditory imagery experiments reported activation peaks within two ALE clusters bilaterally overlapping secondary auditory cortex (planum temporale), indicating reliable activation of these areas during auditory imagery (**Figure 2**). Bilateral activations of inferior frontal cortex were also apparent. Because the imagery tasks used across auditory imagery experiments were non-linguistic in nature (e.g., tone imagery), involvement of Broca's area in auditory imagery was not readily attributable to language-related phonological processing.

MOTOR IMAGERY

Minimum cluster size threshold in the ALE analysis of motor imagery studies was set at $712\,mm^3$. One hundred and fifty

Table 2 | Weighted centers of clusters in the general imagery ALE analysis.

	Region	BA	x	y	z	Volume
All studies	L Superior/Inferior Parietal Lobule/Precuneus	7/40	−28	−56	51	10544
	L Inferior/Middle Frontal Gyrus/Precentral Gyrus	9/47/6	−43	14	18	7216
	R Precuneus/Superior Parietal Lobule	7	22	−63	54	3216
	L Middle Occipital Gyrus	37	−51	−63	−5	2320
	L Middle Frontal Gyrus	46	−41	33	19	2064
	L Middle Frontal Gyrus	6	−28	−1	55	1904
	L Putamen/Caudate/Insula		−25	0	4	1448
	R Medial Frontal Gyrus	6	6	21	45	1376
Conjunction of contrasts vs. rest and non-rest baselines	L Superior/Inferior Parietal Lobule	7/40	−30	−56	52	620
	L Superior Parietal/Precuneus	5/7	−16	−62	54	320
	R Superior Occipital/Parietal Gyrus	7	20	−66	54	150
	R SMA/Med Superior Frontal Gyrus/Cingulum	32/8	6	20	44	90
	L Precentral/Middle Frontal Gyrus	6	−30	0	56	60
	L Inferior Parietal Lobule	40	−38	−38	46	20
	L Precentral/Inferior Frontal Gyrus	44/48	−42	10	28	20
	L Inferior Frontal Gyrus	38/47	−48	24	−6	20
	L Inferior Frontal Gyrus	45	−44	34	18	10

L, Left; R, Right; SMA, Supplementary Motor Area; Med, Medial; BA, Brodmann area. Volume is measured in mm^3. Coordinates reflect standard MNI space.

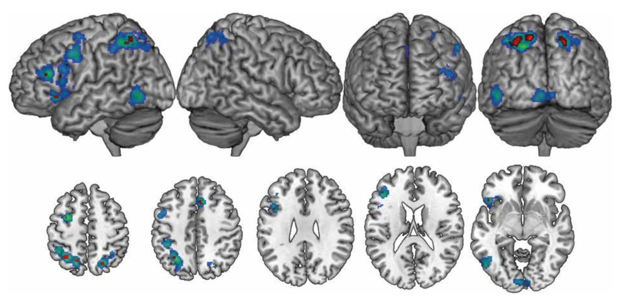


FIGURE 1 | The general imagery network (cool colors) was primarily left-lateralized, with bilateral activations in superior parietal regions. A conjunction analysis of studies employing complex and resting state baselines found nine clusters (red) within the general imagery network that were active across all imagery conditions, regardless of baseline task.

seven foci from 13 experiments involving 137 participants contributed toward these results. For the purposes of this analysis, motor cortex was defined by the Brodmann area 4 definition within the MRIcroN software package. Five clusters were reliably associated with motor imagery at a statistical threshold of $pN < 0.05$ (FDR corrected) (**Table 3**; **Figure 2**). Recruitment of primary motor cortex in motor imagery was not apparent in either hemisphere, though three clusters overlapped to a large extent (right: 222 mm^3; left superior: 608 mm^3; left inferior: 72 mm^3) with premotor cortex. The posterior-most motor imagery cluster, centered at ($x = -37$, $y = -43$, $z = 53$), did overlap substantially with the tactile imagery ROI. The overlapping region was centered at ($x = -38$, $y = -37$, $z = 53$: 1351 mm^3). The 3dClustSim simulations determined that the volume of this overlapping region corresponded to a cluster size corrected threshold of $p < 0.001$ within the primary somatosensory cortex ROI.

TACTILE IMAGERY

Minimum cluster size threshold in the ALE analysis of tactile imagery studies was set at 88 mm^3. Forty-nine foci from four experiments involving 44 participants contributed toward these results. For the purposes of this analysis, primary somatosensory cortex was defined by the union of the Brodmann area 1, 2, and 3 definitions within the MRIcroN software package. Three left-lateralized clusters were reliably associated with motor imagery at a statistical threshold of $pN < 0.05$ (FDR corrected) (**Table 3**; **Figure 2**). Recruitment of primary sensorimotor cortex was apparent in the cluster centered at ($x = -56$, $y = -24$, $z = 43$: 344 mm^3), which overlapped entirely with primary somatosensory cortex. The 3dClustSim simulations determined that the volume of this overlapping region corresponded to a cluster size corrected threshold of $p < 0.001$ within the primary somatosensory cortex ROI. No tactile imagery ALE cluster overlapped with the primary motor cortex ROI.

GUSTATORY IMAGERY

Minimum cluster size threshold in the ALE analysis of gustatory imagery studies was set at 45 mm^3. Note that this cluster size threshold was smaller than the GingerALE-recommended minimum threshold for this dataset. The 3dClustSim analysis determined that 45 mm^3 clusters of size would occur by chance within the gustatory ROI with a probability of 0.05. This reduced cluster threshold permitted the detection of clusters that would reach corrected-level significance in the ROI. Fifty-three foci from five experiments involving 63 participants contributed toward these results. For the purposes of this analysis, gustatory cortex was defined by the AAL template definition

Table 3 | Weighted centers of significant clusters in the auditory, motor, tactile, and gustatory imagery ALE analyses.

Modality	Region	BA	*x*	*y*	*z*	Volume
Auditory	R Superior Temporal Gyrus	22	64	−30	9	2056
	L Inferior Frontal Gyrus	47	−48	24	−5	1360
	L Putamen/Globus Pallidus		−21	−1	4	1136
	L Inferior Frontal Gyrus	44	−51	17	9	1104
	L Superior Temporal Gyrus	22	−60	−38	15	1088
	L Precentral Gyrus	4	−52	1	47	920
	L Inferior Parietal Lobule	40	−58	−38	28	664
	R Inferior Frontal Gyrus	46	56	38	2	648
	L Medial Frontal Gyrus	6	−1	−14	53	640
	L Superior Frontal Gyrus	6	−8	1	69	640
Motor	L Inferior/Superior Parietal Lobule	40/7	−37	−43	53	4464
	L Precentral/Superior Frontal Gyrus	6	−26	−1	56	3584
	R Middle Frontal/Precentral Gyrus	6/4	33	−3	56	1000
	L Inferior Frontal Gyrus	44/45	−57	10	17	976
	L Medial Frontal Gyrus	6	2	5	56	768
Tactile	L Postcentral Gyrus	2	−56	−24	43	344
	L Inferior Frontal Gyrus	46	−51	39	6	96
	L Precentral Gyrus	6	−52	3	50	88
Gustatory	L Middle Frontal Gyrus	46	−41	35	16	272
	L Claustrum		−30	0	12	96
	L Claustrum		−39	0	6	80
	L Precentral Gyrus	6	50	−6	32	80
	R Red Nucleus		3	−27	−15	64
	L Insula		−39	9	3	64
	L Middle Frontal Gyrus	10	−33	43	3	64
	L Precentral Gyrus	6	−50	−4	34	64
	R Middle Frontal Gyrus	6	45	3	51	64
	L Superior Frontal Gyrus	6	−3	15	51	64

L, Left; R, Right; BA, Brodmann area. Volume is measured in mm^3. Coordinates reflect standard MNI space.

of bilateral frontal operculum and anterior bilateral insula ($y > 7$, corresponding to the anterior third of the volume of the AAL template insula definition). The ALE analysis of all gustatory imagery studies found nine clusters that were reliably associated with gustatory imagery (**Table 3**; **Figure 2**). There was evidence for left-lateralized recruitment of gustatory cortex in gustatory imagery: One cluster, centered at (left: $x = -39$, $y = 9$, $z = 3$: $64\,\text{mm}^3$), was overlapped completely by the gustatory cortex definition. The 3dClustSim simulations determined that the volume of this overlapping region corresponded to a cluster size corrected threshold of $p < 0.05$ within the primary gustatory cortex ROI.

OLFACTORY IMAGERY

Minimum cluster size threshold in the ALE analysis of olfactory imagery studies was set at $136\,\text{mm}^3$. Fifty-one foci from five experiments involving 80 participants contributed toward these results. Olfactory cortex was defined by the AAL template definition of bilateral piriform cortex. The ALE analysis of all olfactory imagery studies found four clusters that were reliably associated with olfactory imagery (**Table 4**; **Figure 2**). There was overlap, centered at ($x = -25$, $y = 8$, $z = -16$: $14\,\text{mm}^3$), between olfactory cortex and the third largest cluster centered at ($x = -28$, $y = 11$, $z = -17$). The 3dClustSim simulations determined that the volume of this overlapping region corresponded to a cluster size corrected threshold of $p < 0.001$ within the primary olfactory cortex ROI.

VISUAL IMAGERY

Whether early visual cortex, corresponding to Brodmann areas 17 and 18, participates critically in visual imagery has been a subject of much study (Kosslyn and Thompson, 2003). Visual input is rich in information, however, and we distinguish between different *types* of visual information. Importantly, a number of functionally specialized brain regions contain neurons that are preferentially tuned to different aspects of visual input: The lateral occipital complex (LOC) is specialized for shape processing (Sathian, 2005); neurons in area V4 are tuned to discriminate color (Bramão et al., 2010), and neurons in area V5/MT are critical in the perception of motion (Grèzes, 2001). Kosslyn and Thompson concluded that early visual cortex is involved in visual imagery, in the general sense, when the imagery task requires high-fidelity representations (Kosslyn and Thompson, 2003). An interesting extension to this question is whether the functional organization apparent in visual perception may be found in various subtypes of visual imagery.

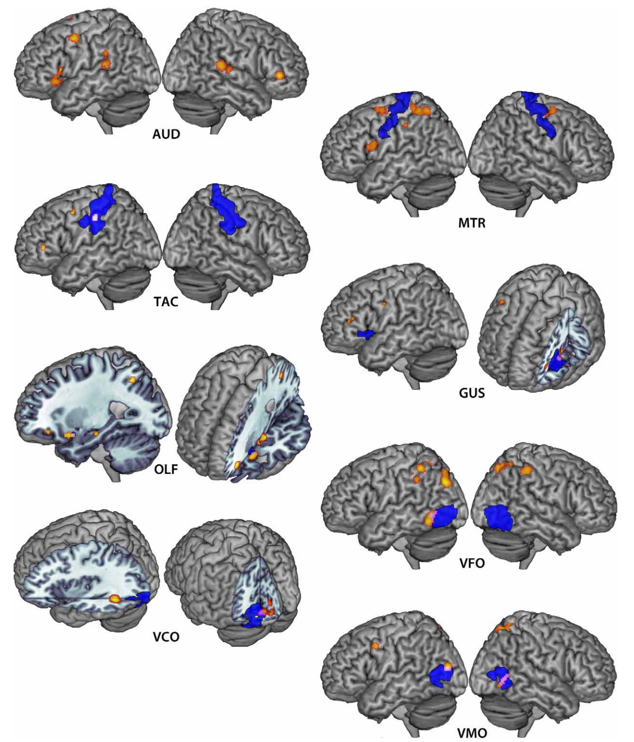


FIGURE 2 | Extent-thresholded clusters of voxels reaching $pN < 0.05$ significance in the auditory (AUD), motor (MTR), tactile (TAC), gustatory (GUS), olfactory (OLF), visual form (VFO), visual color (VCO) and visual motion (VMO) imagery ALE maps are depicted in warm values. Corresponding primary sensorimotor cortices are depicted in violet. Overlapping regions are depicted in pink/white.

This is a strong test of the hypothesis that perceptual processes underlie imagery, as there is no reason that retrieval of stored visual representations—that is, visual information that has already been processed by the perceptual system—should *necessarily* require the involvement of these specialized brain regions. The following three analyses test whether form, color and motion imagery recruits the corresponding functionally specialized visual perception areas. Overlap between the form, color and motion ROIs was avoided by removing voxels appearing in the LOC ROI from the V4 and V5 ROI definitions.

VISUAL FORM IMAGERY

Minimum cluster size threshold in the ALE analysis of visual form imagery studies was set at $2384\,\text{mm}^3$. Two hundred and forty eight foci from 21 experiments involving 218 participants contributed toward these results. For the purposes of this analysis, LOC was defined by the intersection of the Harvard-Oxford Cortical Structural Atlas definition of Lateral Occipital Cortex with the reverse inference map generated by Neurosynth (Yarkoni et al., 2011) for the term "LOC," thresholded at $Z > 5.39$. The ALE analysis of all visual form imagery studies found seven clusters, bilaterally- but primarily left-distributed, reliably associated with visual form imagery (**Table 4**, **Figure 2**). There was overlap, centered at ($x = -52$, $y = -62$, $z = -4$: $631\,\text{mm}^3$), between LOC and the smallest cluster centered at ($x = -55$, $y = -59$, $z = -8$). The 3dClustSim simulations determined that the volume of this overlapping region corresponded to a cluster size corrected threshold of $p < 0.001$ within the LOC ROI. Because the visual ROIs were adjacent to one another, the overlap of the visual form clusters with the color and motion ROIs was additionally assessed. The cluster overlapping with LOC additionally overlapped the visual color ROI definition by $126\,\text{mm}^3$, which corresponded to a cluster size corrected threshold of $p < 0.005$ within the visual color ROI. No visual form ALE cluster overlapped with the visual color ROI.

VISUAL COLOR IMAGERY

Minimum cluster size threshold in the ALE analysis of visual color imagery was set at $192\,\text{mm}^3$. Eighty-one foci from seven experiments involving 76 participants contributed toward these results. For the purposes of this analysis, V4 was defined by the Juelich Histological Atlas definition of left and right V4. The ALE analysis of all visual color imagery studies found 4 left-lateralized clusters that were reliably associated with visual color imagery (**Table 4**, **Figure 2**). Overlap, centered at (left: $x = -18$, $y = -82$, $z = -6$: $42\,\text{mm}^3$), was found between V4 and the largest cluster centered at ($x = -13$, $y = -85$, $z = -4$). The 3dClustSim simulations determined that the volume of this overlapping region corresponded to a cluster size corrected threshold of $p < 0.05$ within the V4 ROI. No visual color ALE cluster overlapped with either the visual form or visual motion ROIs.

VISUAL MOTION IMAGERY

Minimum cluster size threshold in the ALE analysis of visual motion imagery was set at $368\,\text{mm}^3$. One hundred and ten foci from 10 experiments involving 97 participants contributed toward these results. For the purposes of this analysis, V5 was defined by the intersection of the Juelich Histological Atlas definition of left and right V5 with the reverse inference map generated by Neurosynth (Yarkoni et al., 2011) for the term "mt," thresholded at $Z > 5.39$. The ALE analysis of all visual motion imagery studies found six clusters that were reliably associated with visual motion imagery (**Table 4**, **Figure 2**). Bilateral overlap between ALE clusters and V5 was noted (left: $x = -42$, $y = -8$, $z = 14$, $204\,\text{mm}^3$; right: $x = 47$, $y = -61$, $z = 1$, $548\,\text{mm}^3$). The 3dClustSim simulations determined that the volume of both overlapping regions corresponded to a cluster size corrected threshold of $p < 0.001$ within the V5 ROI. The cluster overlapping with right V5 additionally overlapped the visual form ROI definition by $262\,\text{mm}^3$, which corresponded to a cluster size corrected threshold of $p < 0.001$ within the visual form ROI. The form and motion imagery clusters did not overlap within either of the ROIs, however there was a $244\,\text{mm}^3$ overlap between form and motion imagery clusters centered in the right superior parietal lobule ($x = 17$, $y = -66$, $z = 57$; Brodmann area 7). The overlap between form and motion imagery activations in BA 7 is notable in light of the implication of this region in the integration of visual and motor information (Wolpert et al., 1998).

Table 4 | Weighted centers of significant clusters in the olfactory, visual form, visual color, and visual motion imagery ALE analyses.

Modality	Region	BA	x	y	z	Volume
Olfactory	L Anterior Cingulate	32	−22	38	−11	376
	L Hippocampus	28	−23	−18	−17	280
	L Insula/Amygdala	34/38	−28	11	−17	256
	L Superior Parietal Lobule	7	−24	−61	46	256
Vis Form	R Precuneus	7	22	−67	51	3960
	L Inferior/Superior Parietal Lobule/Supramarginal/Angular Gyrus	40	−38	−50	48	3352
	R Lingual Gyrus	17/18	3	−93	−5	2104
	L Superior Occipital Gyrus/Precuneus	39	−33	−78	38	1760
	R Medial Frontal Gyrus	6/32	2	18	47	1424
	R Inferior Parietal Lobule/Supramarginal Gyrus	40	45	−38	46	1400
	L Inferior Temporal Gyrus	20/37	−55	−59	−8	1344
Vis Color	L Lingual Gyrus	18	−13	−85	−4	976
	L Fusiform Gyrus	37	−39	−51	−12	832
Vis Motion	R Precuneus	7	19	−60	61	2120
	R Middle Temporal Gyrus	37	47	−63	0	704
	L Middle Occipital Gyrus	19	−43	−81	16	528
	L Precuneus	31/7	−18	−76	36	528
	L Precuneus	7	−16	−58	59	456
	L Middle Frontal Gyrus	6	−46	7	41	368

L, Left; R, Right; BA, Brodmann area. Volume is measured in mm^3. Coordinates reflect standard MNI space.

GENERAL DISCUSSION

MODALITY-GENERAL IMAGERY

The first goal of this study was to identify the neural substrate underpinning modality-general imagery. Across all sensorimotor modalities, and many experimental paradigms, a core network emerged of brain regions associated with imagery. Activations were seen bilaterally in the general imagery analysis, and in some modalities (auditory, motor, gustatory, visual form and visual motion), but were primarily left-lateralized. It was noted earlier that perceptually-based representational theories assume that multisensory imagery underlies semantic retrieval. Others have suggested that the default-mode network, a well-defined network of brain regions more active during periods of rest than under cognitive load, may arise in part out of introspective processes, including imagery (Daselaar et al., 2010). Though the general imagery network bears a superficial resemblance to the resting state network described in the literature, it does not generally overlap with this network. The imagery network was derived from activations for contrasts of imagery greater than baseline. Activation of the resting state network would thus be precluded by definition. These results should not, therefore, be taken as evidence implying any particular property of the default-mode network. For example, a relative increase in imagery network activation may be apparent when resting state activity is compared to tasks that do not involve imagery.

MODALITY-SPECIFIC IMAGERY

A second goal of this study was to identify the neural substrates underpinning modality-specific imagery, and assess the degree to which imagery in each modality recruited sensorimotor cortex. The ALE analysis of activation loci suggests that modality-specific imagery or knowledge retrieval for most modalities is associated with increased activation in corresponding sensorimotor regions. Though modalities differ with respect to the lateralization and extent of this recruitment, this suggests that modality-specific imagery generally recruits the corresponding primary perceptual areas. Whether these differences reflect differences in cognitive processing, or have behavioral implications remains unclear. For example, proportionally greater recruitment of perceptual regions may be associated with higher fidelity imagery, whereas greater recruitment of adjacent areas is associated with more abstract (e.g., linguistically-dependent) manipulations of imagery representations.

One challenge for this interpretation concerns the failure to show recruitment of primary sensorimotor perceptual cortices for the auditory and motor modalities. The ALE analyses showed imagery in these modalities does reliably recruit posterior superior temporal gyrus (STG) and premotor cortex, respectively. These results are consistent with Kosslyn et al. (2001) review finding that auditory imagery does not activate primary auditory cortex (A1), but does activate auditory associative areas. The same review concluded that motor imagery conditionally activates motor areas, but required a more liberal definition of motor area: Of the studies reviewed, most reported imagery-related activations in premotor cortex but not primary motor cortex. Posterior STG and premotor cortex have been associated with maintaining auditory and motor sequence representations, respectively (Ohbayashi et al., 2003; Arnott et al., 2005; Buchsbaum and D'Esposito, 2008). Thus, an alternative interpretation of imagery-related activations is that they reflect activations within memory systems for these modalities, and that these systems are situated adjacent to, rather than within primary auditory and motor cortices. This may indeed be the case, though such a conclusion rests on the sort of circular logic that highlights the centrality of the symbol grounding problem to understanding the neural bases of cognitive processes. It thus remains to be seen whether a satisfactory solution to the symbol grounding problem can be found for these imagery-related processes. These patterns are, however, suggestive of a modality-specific working memory system.

Though the question of whether visual imagery, in the general sense, recruits early visual cortex has been extensively studied (Kosslyn and Thompson, 2003), the more specific question of whether the functional distinction of color, motion and form perception is reliably found in visual imagery has remained unclear. A third goal of the present study was to determine whether similar functional specialization occurs in visual imagery. The present results indicate that, though visual imagery may activate early visual areas, imaginary color, motion, and shape processing is facilitated by upstream visual areas specialized for color, motion, and form perception, respectively. This parallel specialization during visual imagery is interesting in light of the fact that imagery involves the retrieval of stored representations. That is, imagery is based on information previously processed by the perceptual system. Nonetheless, imagery recruits brain regions involved in processing the original perceptual stream. To retrieve pre-processed rather than post-processed representations would thus be a sub-optimal strategy unless it conveys some other benefit. One possibility is that this processing reflect does not reflect processing of the raw visual stream. Rather, these regions may encode perceptual patterns that are reliably associated with information in other modalities. If this information is captured in the perceptual processing stream, it would be unnecessary to encode this information at higher levels of abstraction. Thus, imagery processes implied by perceptually-based representational theories may recruit these areas in order to generate more veridical multisensory representations.

Several crossmodal asymmetries were observed within the modality-specific imagery results. When the results of the motor and tactile imagery analyses are taken together, they suggest that motor imagery may imply a tactile component, but not the converse. This asymmetry may arise from the types of motor imagery tasks used: in more than half of the motor imagery tasks, the task implied imagery of an action on an object. This asymmetry would be predicted by the dependency of imagery on perceptual experience: one is often passively touched by objects (i.e., tactile perception without an associated motor response), but seldom acts on an object without also touching it. Similarly, an asymmetrical relationship existed among the three visual modality subtypes: First, form imagery clusters additionally overlapped the color ROI, but not *vice versa*. Second, motion imagery clusters additionally overlapped the form ROI, but not *vice versa*. This second asymmetry plausibly reflects our visual experience of moving objects: Form processing may be commonly implicated in motion processing because one typically perceives motion of an object with form. The converse relationship does not seem quite as strong, as we regularly encounter inanimate forms that do not move. In contrast, the apparently consistent recruitment of primary color processing regions during form imagery, but not the converse, is puzzling. We do not usually experience fields of color, but instead see colored objects, or *forms*. On the other hand, we do regularly experience well-defined forms without any associated color: square vs. oval windows, for example. The observed asymmetry would thus appear to be the reverse of what one would expect on the basis of real-world experience. One possibility is that it reflects an interaction between a statistical artifact of the number of form imagery studies and the proximity of the two regions of interest. More visual form imagery studies were conducted with more participants, generating more extensive ALE maps, with a higher probability of overlapping an adjacent ROI. Alternatively, it may reflect a real property of the systems involved in color and form imagery, though that remains a subject for future investigation.

MODALITY-SPECIFIC IMAGERY AND PERCEPTUALLY-GROUNDED REPRESENTATIONS

Finally, and perhaps most importantly for investigations of perceptually-grounded representations, in no modality were imagery clusters restricted to brain regions immediately involved in perception. Those clusters that did overlap with primary somatosensory regions generally extended beyond these areas. In contrast to perception or imagery-based accounts of knowledge representations, amodal models of semantic memory assume concept knowledge is maintained as an abstraction bearing no connection to perceptual processing (Pylyshyn, 1973; Tyler and Moss, 2001). It is no less reasonable to suppose that a modality-specific representational system encodes information in sensory association areas, but not necessarily in primary sensorimotor areas. This perspective is consistent with Thompson-Schill's review of neuroimaging studies of semantic memory (Thompson-Schill, 2003), which concluded that the literature supported a distributed modality-specific semantic system, but that "studies which have directly compared semantic retrieval and perception have consistently found an anterior shift in activation during semantic processing" (p. 283).

The present meta-analysis suggests that a perceptually-grounded representational system recruits primary sensory cortex to a modest and varying degree, but that processing relies greatly on upstream (though not necessarily anatomically anterior) unimodal convergence zones (Binder and Desai, 2011; McNorgan et al., 2011). These regions tend to be adjacent to their associated perceptual areas, and integrate downstream perceptual codes into somewhat more abstract (but perceptually-grounded) representations. This account would be consistent with the distribution of modality-specific imagery activations about primary somatosensory cortices, and with the theoretical ties between modality-specific representations and imagery. This interpretation would also be consistent with a recent investigation of visual imagery and memory by Slotnick et al. (2011) in which the authors concluded that "visual memory and visual mental imagery are mediated by largely overlapping neural substrates in both frontal-parietal control regions and occipital-temporal sensory regions" (p. 20). These results suggest that neuroimaging investigations of perceptually-based knowledge might pay particular attention to primary sensorimotor areas also implied in imagery, but also should consider contributions of other brain regions supporting imagery processes.

CONCLUSIONS

Though neuroscientific studies of imagery have proliferated over the last decade, not all forms of imagery have been investigated to the same extent—imagery of the chemical senses and tactile imagery appear to be relatively underrepresented. Some of the questions posed here may not be adequately answerable without further study in these imagery domains. Similarly, the present review omits studies of imagery in other domains, such as emotional, temporal or spatial imagery, which may be more abstract forms of meta-imagery involving the integration of multiple modalities or function as representational primitives.

Finally, these results are generally consistent with the assumption that mental imagery underlies representational knowledge, though the matter is far from resolved. These considerations point toward a need for further investigation in the imagery domain. These efforts will help relate cognitive processes to one another and to arrive at a fully grounded model of cognitive processing.

REFERENCES

Alivisatos, B., and Petrides, M. (1997). Functional activation of the human brain during mental rotation. *Neuropsychologia* 35, 111–118.

Arnott, S. R., Grady, C. L., Hevenor, S. J., Graham, S., and Alain, C. (2005). The functional organization of auditory working memory as revealed by fMRI. *J. Cogn. Neurosci.* 17, 819–831.

Barnes, J., Howard, R. J., Senior, C., Brammer, M., Bullmore, E. T., Simmons, A., et al. (2000). Cortical activity during rotational and linear transformations. *Neuropsychologia* 38, 1148–1156.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Beauchamp, M. S., Haxby, J. V., Jennings, J. E., and Deyoe, E. A. (1999). An fMRI version of the farnsworth–munsell 100-Hue test reveals multiple color-selective areas in human ventral occipitotemporal cortex. *Cereb. Cortex* 9, 257–263.

Belardinelli, M. O., Palmiero, M., Sestieri, C., Nardo, D., Di Matteo, R., Londei, A., et al. (2009). An fMRI investigation on image generation in different sensory modalities: the influence of vividness. *Acta Psychol.* 132, 190–200.

Binder, J. R., and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn. Sci.* 15, 527–536.

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., et al. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language A positron emission tomography activation study. *Brain* 117, 1241–1253.

Bramão, I., Faísca, L., Forkstam, C., Reis, A., and Petersson, K. M. (2010). Cortical brain regions associated with color processing: an FMRi Study. *Open Neuroimaging J.* 4, 164–173.

Buchsbaum, B. R., and D'Esposito, M. (2008). The Search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20, 762–778.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H.-J., and Jancke, L. (2005). Scanning silence: mental imagery of complex sounds. *Neuroimage* 26, 1119–1127.

Canessa, N., Borgo, F., Cappa, S. F., Perani, D., Falini, A., Buccino, G., et al. (2007). The different neural correlates of action and functional knowledge in semantic memory: an fMRI study. *Cereb. Cortex* 18, 740–751.

Chein, J. M., Fissell, K., Jacobs, S., and Fiez, J. A. (2002). Functional heterogeneity within Broca's area during verbal working memory. *Physiol. Behav.* 77, 635–639.

Chouinard, P. A., and Goodale, M. A. (2010). Category-specific neural processing for naming pictures of animals and naming pictures of tools: an ALE meta-analysis. *Neuropsychologia* 48, 409–418.

Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684.

Creem-Regehr, S. H., Neil, J. A., and Yeh, H. J. (2007). Neural correlates of two imagined egocentric transformations. *Neuroimage* 35, 916–927.

Daselaar, S. M., Porat, Y., Huijbers, W., and Pennartz, C. M. A. (2010). Modality-specific and modality-independent components of the human imagery system. *Neuroimage* 52, 677–685.

Dechent, P. (2004). Is the human primary motor cortex involved in motor imagery? *Cogn. Brain Res.* 19, 138–144.

De Lange, F. P., Hagoort, P., and Toni, I. (2005). Neural topography and content of movement representations. *J. Cogn. Neurosci.* 17, 97–112.

D'Esposito, M., Detre, J. A., Aguirre, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J., et al. (1997). A functional MRI study of mental image generation. *Neuropsychologia* 35, 725–730.

Djordjevic, J., Zatorre, R. J., Petrides, M., Boyle, J. A., and Jones-Gotman, M. (2005). Functional neuroimaging of odor imagery. *Neuroimage* 24, 791–801.

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., and Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *Neuroimage* 59, 2349–2361.

Farnsworth, D. (1957). *The Farnsworth-Munsell 100-Hue Test for the Examination of Color Vision.* Baltimore, MD: Munsell Color Company.

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J., and Dolan, R. J. (1996). The trouble with cognitive subtraction. *Neuroimage* 4, 97–104.

Ganis, G. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cogn. Brain Res.* 20, 226–241.

Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., and Singer, W. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur. J. Neurosci.* 10, 1563–1573.

Gottfried, J. A., and Dolan, R. J. (2004). Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nat. Neurosci.* 7, 1144–1152.

Grèzes, J. (2001). Does perception of biological motion rely on specific brain regions? *Neuroimage* 13, 775–785.

Guillot, A., Collet, C., Nguyen, V. A., Malouin, F., Richards, C., and Doyon, J. (2009). Brain activity during visual versus kinesthetic imagery: an fMRI study. *Hum. Brain Mapp.* 30, 2157–2172.

Gulyás, B. (2001). Neural networks for internal reading and visual imagery of reading: a PET study. *Brain Res. Bull.* 54, 319–328.

Halpern, A. R., and Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb. Cortex* 9, 697–704.

Hanakawa, T. (2002). Functional properties of brain areas associated with motor execution and imagery. *J. Neurophysiol.* 89, 989–1002.

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Howard, R. J., Ffytche, D. H., Barnes, J., Mckeefry, D., Ha, Y., Woodruff, P. W., et al. (1998). The functional anatomy of imagining and perceiving colour. *Neuroreport* 9, 1019–1023.

Hsu, N. S., Frankland, S. M., and Thompson-Schill, S. L. (2012). Chromaticity of color perception and object color knowledge. *Neuropsychologia* 50, 327–333.

Hsu, N. S., Kraemer, D. J. M., Oliver, R. T., Schlichting, M. L., and Thompson-Schill, S. L. (2011). Color, context, and cognitive style: variations in color knowledge retrieval as a function of task and subject variables. *J. Cogn. Neurosci.* 23, 2544–2557.

Ishai, A., Ungerleider, L. G., and Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron* 28, 979–990.

Johnson, S. (2002). Selective activation of a parietofrontal circuit during implicitly imagined prehension. *Neuroimage* 17, 1693–1704.

Jordan, K., Heinze, H. J., Lutz, K., Kanowski, M., and Jancke, L. (2001). Cortical activations during the mental rotation of different visual objects. *Neuroimage* 13, 143–152.

Kaas, A., Weigelt, S., Roebroeck, A., Kohler, A., and Muckli, L. (2010). Imagery of a moving object: the role of occipital cortex and human MT/V5+. *Neuroimage* 49, 794–804.

Kellenbach, M. L., Brett, M., and Patterson, K. (2001). Large, colorful, or noisy? Attribute- and modality-specific activations during retrieval of perceptual attribute knowledge. *Cogn. Affect. Behav. Neurosci.* 1, 207–221.

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., and Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *J. Neurosci.* 28, 12224–12230.

Kikuchi, S., Kubota, F., Nisijima, K., Washiya, S., and Kato, S. (2005). Cerebral activation focusing on strong tasting food: a functional magnetic resonance imaging study. *Neuroreport* 18, 281–283.

Kobayashi, M., Takeda, M., Hattori, N., Fukunaga, M., Sasabe, T., Inoue, N., et al. (2004). Functional imaging of gustatory perception and imagery: top-down processing of gustatory signals. *Neuroimage* 23, 1271–1282.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., et al. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *J. Cogn. Neurosci.* 5, 263–287.

Kosslyn, S. M., Ganis, G., and Thompson, W. L. (2001). Neural foundations of imagery. *Nat. Rev. Neurosci.* 2, 635–642.

Kosslyn, S. M., and Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychol. Bull.* 129, 723–746.

Kosslyn, S. M., Thompson, W. L., and Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: a positron emission tomography study. *Neuroimage* 6, 320–334.

Kosslyn, S. M., Thompson, W. L., Klm, I. J., and Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature* 378, 496–498.

Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., et al. (2005). ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. *Hum. Brain Mapp.* 25, 155–164.

McNorgan, C., Reid, J., and McRae, K. (2011). Integrating conceptual knowledge within and across representational modalities. *Cognition* 118, 211–233.

Mellet, E., Tzourio, N., Crivello, F., Joliot, M., Denis, M., and Mazoyer, B. (1996). Functional anatomy of spatial mental imagery generated from verbal instructions. *J. Neurosci.* 16, 6504–6512.

Newman, S. D., Klatzky, R. L., Lederman, S. J., and Just, M. A. (2005). Imagining material versus geometric properties of objects: an fMRI study. *Cogn. Brain Res.* 23, 235–246.

Nyberg, L. (2001). Reactivation of motor brain areas during explicit memory for actions. *Neuroimage* 14, 521–528.

Nyberg, L., Habib, R., McIntosh, A. R., and Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11120–11124.

Ohbayashi, M., Ohki, K., and Miyashita, Y. (2003). Conversion of working memory to motor sequence in the monkey premotor cortex. *Science* 301, 233–236.

Oliver, R. T., Geiger, E. J., Lewandowski, B. C., and Thompson-Schill, S. L. (2009). Remembrance of things touched: how sensorimotor experience affects the neural instantiation of object form. *Neuropsychologia* 47, 239–247.

Paivio, A. (1971). *Imagery and Verbal Processes.* New York, NY: Holt, Rinehart and Winston.

Plailly, J., Delon-Martin, C., and Royet, J.-P. (2012). Experience induces functional reorganization in brain regions involved in odor imagery in perfumers. *Hum. Brain Mapp.* 33, 224–234.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* 10, 59–63.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychol. Bull. Psychol. Bull.* 80, 1–24.

Roland, P. E., and Gulyás, B. (1995). Visual memory, visual imagery, and visual recognition of large field patterns by the human brain: functional anatomy by positron emission tomography. *Cereb. Cortex* 5, 79–93.

Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550.

Sack, A. T., Sperling, J. M., Prvulovic, D., Formisano, E., Goebel, R., Di Salle, F., et al. (2002). Tracking the mind's image in the brain II: transcranial magnetic stimulation reveals parietal asymmetry in visuospatial imagery. *Neuron* 35, 195–204.

Sathian, K. (2005). Visual cortical activity during tactile perception in the sighted and the visually deprived. *Dev. Psychobiol.* 46, 279–286.

Servos, P., Osu, R., Santi, A., and Kawato, M. (2002). The neural substrates of biological motion perception: an fMRI study. *Cereb. Cortex* 12, 772–782.

Simmons, W. K., and Barsalou, L. W. (2003). The similarity-in-topography principle: reconciling theories of conceptual deficits. *Cogn. Neuropsychol.* 20, 451–486.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., and Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia* 45, 2802–2810.

Slotnick, S. D., Thompson, W. L., and Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cereb. Cortex* 15, 1570–1583.

Slotnick, S. D., Thompson, W. L., and Kosslyn, S. M. (2011). Visual memory and visual mental imagery recruit common control and sensory regions of the brain. *Cogn. Neurosci.* 3, 14–20.

Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M., and Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* 39, 701–711.

Thompson, W. L., Kosslyn, S. M., Sukel, K. E., and Alpert, N. M. (2001). Mental imagery of high- and low-resolution gratings activates area 17. *Neuroimage* 14, 454–464.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: inferring "how" from "where". *Neuropsychologia* 41, 280–292.

Trojano, L., Grossi, D., Linden, D. E., Formisano, E., Hacker, H., Zanella, F. E., et al. (2000). Matching two imagined clocks: the functional anatomy of spatial analysis in the absence of visual stimulation. *Cereb. Cortex* 10, 473–481.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., and Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage* 16, 765–780.

Tyler, L. K., and Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends Cogn. Sci.* 5, 244–252.

Veldhuizen, M. G., Bender, G., Constable, R. T., and Small, D. M. (2007). Trying to detect taste in a tasteless solution: modulation of early gustatory cortex by attention to taste. *Chem. Senses* 32, 569–581.

Vingerhoets, G., Delange, F., Vandemaele, P., Deblaere, K., and Achten, E. (2002). Motor imagery in mental rotation: an fMRI study. *Neuroimage* 17, 1623–1633.

Wager, T. D., Lindquist, M., and Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: current and future directions. *Soc. Cogn. Affect. Neurosci.* 2, 150–158.

Warrington, E. K., and McCarthy, R. A. (1987). Categories of knowledge. Further fractionations and an attempted integration. *Brain* 110(Pt 5), 1273–1296.

Wheeler, M. E., Petersen, S. E., and Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11125–11129.

Wolpert, D. M., Goodbody, S. J., and Husain, M. (1998). Maintaining internal representations: the role of the human superior parietal lobe. *Nat. Neurosci.* 1, 529–533.

Yágüez, L., Nagel, D., Hoffman, H., Canavan, A. G. M., Wist, E., and Hömberg, V. (1998). A mental route to motor learning: improving trajectorial kinematics through imagery training. *Behav. Brain Res.* 90, 95–106.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Meth.* 8, 665–670.

Yeshurun, Y., Lapid, H., Dudai, Y., and Sobel, N. (2009). The privileged brain representation of first olfactory associations. *Curr. Biol.* 19, 1869–1874.

Yomogida, Y. (2004). Mental visual synthesis is originated in the fronto-temporal network of the left hemisphere. *Cereb. Cortex* 14, 1376–1383.

Yoo, S.-S., Freeman, D. K., McCarthy, J. J. I., and Jolesz, F. A. (2003). Neural substrates of tactile imagery: a functional MRI study. *Neuroreport* 14, 581–585.

Yoo, S.-S., Kee, C. U., and Choi, B. G. (2001). Human brain mapping of auditory imagery: event-related functional MRI study. *Neuroreport* 12, 3045–3049.

Zatorre, R. J., and Halpern, A. R. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. *J. Cogn. Neurosci.* 8, 29–46.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 August 2012; accepted: 28 September 2012; published online: 17 October 2012.

Citation: McNorgan C. (2012) A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. Front. Hum. Neurosci. ***6****:285. doi: 10.3389/fnhum.2012.00285*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 31 December 2012
doi: 10.3389/fnhum.2012.00333

Conceptual structure within and between modalities

Katia Dilkina and Matthew A. Lambon Ralph*

Neuroscience and Aphasia Research Unit, School of Psychological Sciences, University of Manchester, Manchester, UK

Edited by:
Paul D. Siakaluk, University of Northern British Columbia, Canada

Reviewed by:
Jamie Reilly, University of Florida, USA
Sebastian Crutch, University College London, UK

****Correspondence:***
Matthew A. Lambon Ralph, Neuroscience and Aphasia Research Unit, School of Psychological Sciences, University of Manchester, Zochonis Building, Oxford Road, Manchester M13 9PL, UK.
e-mail: matt.lambon-ralph@manchester.ac.uk

Current views of semantic memory share the assumption that conceptual representations are based on multimodal experience, which activates distinct modality-specific brain regions. This proposition is widely accepted, yet little is known about how each modality contributes to conceptual knowledge and how the structure of this contribution varies across these multiple information sources. We used verbal feature lists, features from drawings, and verbal co-occurrence statistics from latent semantic analysis to examine the informational structure in four domains of knowledge: perceptual, functional, encyclopedic, and verbal. The goals of the analysis were three-fold: (1) to assess the structure within individual modalities; (2) to compare structures between modalities; and (3) to assess the degree to which concepts organize categorically or randomly. Our results indicated significant and unique structure in all four modalities: perceptually, concepts organize based on prominent features such as shape, size, color, and parts; functionally, they group based on use and interaction; encyclopedically, they arrange based on commonality in location or behavior; and verbally, they group associatively or relationally. Visual/perceptual knowledge gives rise to the strongest hierarchical organization and is closest to classic taxonomic structure. Information is organized somewhat similarly in the perceptual and encyclopedic domains, which differs significantly from the structure in the functional and verbal domains. Notably, the verbal modality has the most unique organization, which is not at all categorical but also not random. The idiosyncrasy and complexity of conceptual structure across modalities raise the question of how all of these modality-specific experiences are fused together into coherent, multifaceted yet unified concepts. Accordingly, both methodological and theoretical implications of the present findings are discussed.

Keywords: concepts, conceptual organization, semantic system, modality-specific knowledge, multimodal knowledge, hub-and-spoke model

INTRODUCTION

We experience objects and entities in the world through many different modalities. We experience them perceptually through observation (visual, auditory, olfactory, gustatory, and tactile), as well as functionally through interaction and use. Remember the eggs that you had for breakfast this morning? Consider the richness of this simple perceptual and motor experience. In addition, we experience objects contextually or in relation to other objects, entities, or places. For example, eggs are usually eaten at breakfast, often accompanied by sausage or bacon. Finally, there is the abundant verbal experience when we read, write, or talk about things in the world.

Each of these modalities provides a rich and unique experience, and contributes to our semantic knowledge – our cross-modal conceptual knowledge. All contemporary theories of semantic memory and its neural basis share the assumption that semantic representations are formed from multimodal experience, coded in distinct modality-specific brain regions. Furthermore, the regions representing information relevant to a specific item are activated during semantic processing whether or not this type of information is explicitly required by the task/activity. Considerable evidence has been accumulated in favor of these ideas (see Martin, 2007; Patterson et al., 2007; Thompson-Schill, 2003 for reviews). For example, in a PET neuroimaging paradigm, Martin et al. (1995) asked a group of participants to name either the color or the action associated with visually presented objects. They found that generating action words activated the left posterior middle temporal gyrus (associated with visual motion processing), while producing color words activated the fusiform gyrus (associated with visual form and color processing). Similar patterns of activation were obtained in different experimental paradigms which did not explicitly present or require object properties of specific knowledge type, including brief picture viewing (Chao et al., 1999), picture naming (e.g., Martin et al., 1996; Chao et al., 1999; Moore and Price, 1999), visual match-to-sample (Chao et al., 1999), and same/different judgments with pairs of pictures or words (Perani et al., 1999). In all of these studies, manipulable objects such as tools tend to activate the left posterior MTG more than living things such as animals, which preferentially engage posterior inferior temporo-occipital regions. More detailed investigations have identified distinct areas in the left posterior lateral temporal lobe responding to biological motion vs. object motion (e.g., Beauchamp et al., 2002). The distributed nature of conceptual representations has been investigated not only in the visual and motor modalities, but also in other perceptual modalities. For example, Simmons et al. (2005) presented their participants with pictures of appetizing foods along with pictures of locations. They used a simple same/different judgment task to elicit fMRI activation contrasts for these two categories. The results showed that even in this simple task, food items preferentially activated two areas associated with gustatory/olfactory processing – the right insula and the left orbitofrontal cortex (see also Goldberg et al., 2006; González et al., 2006). These neuroimaging findings have been complemented by neuropsychological reports of patients with category-specific semantic deficits (greater impairment for living vs. non-living things or vice versa, greater impairment for fruits and vegetables vs. tools, and so on), where lesion locations match the brain areas and functional specialization suggested by the functional imaging studies with unimpaired individuals (e.g., Tranel et al., 1997).

Another group of related studies have explored the relationship between conceptual categories and different types of modality-specific information sources by analyzing verbally generated feature lists from unimpaired individuals. For example, Garrard et al. (2001) collected directed feature norms for 62 items from six categories. They classified the features as sensory, functional, or encyclopedic (see Materials and Methods) and investigated how the three types of features compared across the six categories. They found that living things tended to have more sensory than functional attributes compared to non-living things; they also had more encyclopedic attributes. Furthermore, living things had less distinctive and more shared features than non-living things. A similar investigation, though on a substantially larger scale, was conducted by Cree and McRae (2003), who collected verbal feature norms for 541 concrete concepts. The authors argued that a full understanding of category-specific deficits would have to go beyond the distinction between sensory and functional features. They proposed a classification consisting of nine different knowledge types processed in distinct neural regions (**Table 1**). Using their detailed classification, Cree and McRae (2003) showed that all feature types play a role in distinguishing among conceptual categories, though admittedly some knowledge types were more relevant to specific categories than others. For example, not surprisingly, the feature type of visual motion was important for the category of creatures but unimportant for fruits and vegetables, which relied more on visual-color features, etc. The results from a large hierarchical cluster analysis combining the knowledge types with a number of other factors including the proportion of distinguishing features for each category, visual similarity and complexity, semantic similarity, concept familiarity, and word frequency, demonstrated a unique and significant contribution of all factors to the conceptual structure present in their data set. In order to go beyond just those feature types that are most readily reported verbally, Hoffman and Lambon Ralph (2012) asked participants to rate the importance of each sensory and verbal modality to 100 different concepts. Not only did the results complement the previous verbal feature listings but the study found that information arising in other modalities such as sound, motion, smell, and taste (which are rarely reported in verbal listing studies) also provided important differential experience across categories.

In summary, previous investigations have focused on the relationship between categories and feature types. While these studies have provided important insights about semantic representation and the nature of category-specific deficits, in this investigation we returned to consider the primary hypothesis held by contemporary theories of semantic memory – namely that concepts arise from the convergence of our multimodal, verbal, and non-verbal experience. As such, it becomes important to understand the distribution and structure of representation in each modality and how this contributes to the overall multimodal semantic representation. Accordingly, a series of fundamental questions arise: What does the structure within each of these modalities look like? Is it random? Is it categorical/taxonomic? What principles govern the organization of information within modality? How do they compare across modalities?

The goal of this study was to investigate the formation of concepts overall. Thus, we took a novel approach in order to look in more detail at four different modalities of knowledge – visual, verbal, encyclopedic, and functional. We investigated how each modality contributes to the overall semantic representation and how the structure in each modality varies.

There are different ways in which each modality can be probed empirically and how the data arising are treated. We, therefore, compared across the methods directly. As far as we are aware, these direct comparisons have not been made before. For example, we compared information about visual experience as derived from verbal feature listings (two methods each deriving features in a slightly different way) as well as those extracted from participants' drawings of the same concepts. Secondly, we compared feature data (itself a reflection of verbal experience) against the structure present in a very large verbal corpus which does not attempt to derive attributes but instead utilizes co-occurrence statistics to infer the underlying representations [in this case, derived from latent semantic analysis (LSA) of the British National Corpus].

Our study had three distinct goals: (1) to establish the organization of information arising in each modality of knowledge; (2) to compare the structure between the various modalities; and (3) to assess the degree to which concepts in each modality are organized taxonomically or randomly.

Table 1 | Types of knowledge and associated brain regions assumed by Cree and McRae (2003).

Knowledge type	Brain region
Visual – color	Bilateral posterior ventral temporal cortex
Visual – parts and surface properties	Bilateral ventral occipital cortex
Visual – motion	Left posterior middle temporal gyri
Tactile	Dominant hand and finger areas of primary somatosensory and motor cortices
Olfactory	Piriform cortex and right lateral orbitofrontal cortex
Gustatory	(Left) anterior insula and orbitofrontal and precentral gyri† (Kobayashi, 2006)
Auditory	Bilateral posterior superior temporal gyrus†
Functional	Left ventral premotor cortex
Encyclopedic	Multiple regions

† Area not specified by Cree and McRae (2003).

MATERIALS AND METHODS

We utilized four different data sets reported in the literature, which gave us four different types of representations – perceptual, functional, encyclopedic, and verbal. Each of the data sets is described below.

GARRARD ET AL. FEATURE LISTINGS DATA SET

Garrard et al. (2001) asked 20 adult volunteers (mean age 67 years old) to list features for each of 62 items (originally 64 but two of the items were subsequently excluded). The participants were prompted to list the category of each item as well as two to six descriptive features ("an elephant is. . ."), two to six parts features ("an elephant has. . ."), and two to six abilities-and-uses features ("an elephant can. . ."). After the initial data collection, the features were processed to use standardized wording (since a given feature can often be described in multiple ways) and to exclude qualifying, exemplifying, or highly idiosyncratic information. Only features listed by at least two participants were considered. They were classified as sensory, functional (action, activity, or use of an item), encyclopedic (associative relationships), or categorical. The set of features consisted of 50% sensory, 28% functional, 15% encyclopedic, and 7% categorical. We used the sensory, functional, and encyclopedic features from this set.

CREE AND McRae FEATURE LISTINGS DATA SET

Cree and McRae (2003) reported a list of features produced by undergraduate students for a set of 541 concepts commonly used in categorization and semantic memory tasks. In their feature elicitation method, each participant was presented with 20 or 24 (mostly dissimilar) concept names alongside 10 blank lines to be filled with features of each item. Thirty participants provided features for each of the 541 concepts. Only features that were listed by at least five participants were included in the report. The feature listings were not edited in any way. Excluding taxonomic labels, the authors classified each of the features as belonging to one of nine types: seven perceptual types (visual – color, visual – parts, and surface properties, visual – motion, smell, taste, touch, and sound), a functional type (how one interacts with the item), and an encyclopedic type (all non-perceptual and non-functional descriptors). They reasoned that these were widely accepted knowledge types, which are processed in distinct neural regions (see **Table 1**).

Since most of the perceptual knowledge types included few features, we combined them into a single perceptual classification, along with the functional and encyclopedic feature types – a classification comparable to the one used by Garrard et al. (2001).

ROGERS ET AL. PICTURE DRAWINGS DATA SET

Instead of using verbal feature listings, Rogers et al. (2004) asked eight participants (mean age of 62) to draw from name the same 64 items as Garrard et al. (2001) had presented to their participants for verbal feature generation. The subjects had 1 min to draw each item and were told that their drawings would not be judged for artistic merit but be assessed for the degree to which they correctly represented the nature of the object. Two independent raters compiled lists of all the visual features present each drawing. Features included by only a single participant were excluded. After this initial data collection, the feature lists were compared to the drawings once more and features that described overlapping or similar aspects of the drawings were combined together to produce the final visual feature description of each item. We considered the Rogers et al. set as another source of information about how perceptual experience contributes to our conceptual knowledge and compared this information source to the subset of verbal features given a perceptual classification (see above).

HOFFMAN ET AL. LATENT SEMANTIC ANALYSIS DATA SET

The final data set we utilized did not include feature lists. Instead, it provided representations for numerous items based on their patterns of occurrence in verbal context. Hoffman et al. (2011, 2012) performed LSA on the British National Corpus, which contains over 87 million words in 3125 documents. The authors split the original documents into 1000-word-long samples, which gave them 87,375 smaller documents. Only words that appeared at least 50 times in the corpus and in at least 40 different documents were included. The resultant LSA produced 300-dimensional representational vectors for 38,456 words[1].

In order to compare the representations across the various knowledge types, we took the intersection of these data sets, which gave us a list of 52 concepts. **Table 2** presents basic statistics for each of the feature data sets. Two things should be noted. First, even though the 52 items were present in each data set, they did not necessarily have entries for all feature types: 47 items had encyclopedic features in the Garrard et al. set; another subset of 47 had encyclopedic features in the Cree and McRae set; and a smaller subset of 38 had functional features in the Cree and McRae set. Secondly, an important aspect of the feature-based representations is their density. While all three data sets included a great number of features, the percent of features per item, that is the average number of features listed for a single concept divided by the total number of features of this type (i.e., representational density), was relatively low. Most strikingly, the Cree and McRae data set, which was the only one where the initially collected lists of features were not further processed or edited, had the lowest density, for all feature types.

Why might low representational density indicate a problem? When there is a great number of features but each of these features applies to singular or few items, it is possible that there is some information missing – that some of these features in fact apply to more of the items. The problem is more severe than simply that of "missing" information because, in a binary type of representation, each item either has a feature or it does not; there is no such thing as "unknown." So when a feature is missing, it is effectively non-existent for the item. For example, when the data set fails to specify that a dog has a neck, it in fact specifies that a dog does not have a neck.

[1]The high dimensional semantic space resulting from Latent Semantic Analysis is a mathematical representation of a large set of terms (words, phrases) and it is unique to the corpus used. In other words, for the selected corpus and restrictions applied (frequency of occurrence, part of speech, etc.), LSA represents each term as a vector. This vector has no meaning, other than in relation to other vectors in the same semantic space (that is, other terms from the corpus). The relationship between two terms can be quantified with any distance or similarity measure applied to the pair of representational vectors, such as the Euclidean distance or the cosine of the angle between the vectors.

Table 2 | Descriptive statistics for each of the data sets.

Data set	*N* concepts	*N* features	Average *N* concepts per feature	Average *N* features per concept	Density of representation (% feats/concept)
Rogers et al. visual	52	194	3.89	14.50	7.5
Garrard et al. sensory	52	206	3.00	11.87	5.8
Garrard et al. functional	52	144	2.27	6.29	4.4
Garrard et al. encyclopedic	47	79	2.23	3.74	4.7
Cree and McRae perceptual	52	208	2.09	8.35	4.0
Cree and McRae functional	38	87	1.32	3.03	3.5
Cree and McRae encyclopedic	47	112	1.39	3.21	3.0
Cree and McRae edited perceptual	52	214	4.22	17.37	8.1
Cree and McRae edited functional	52	81	2.68	4.17	5.2
Cree and McRae edited encyclopedic	52	115	2.86	6.33	5.5

Since we had multiple sets providing overlapping categories of features, it became clear that one set (Cree and McRae) repeatedly exhibited lower representational density compared to the corresponding sets from alternative sources. To ensure that the low representational density of this data set did not indicate the problem described above, we edited the feature list for each concept in a fashion similar to Rogers et al. (2004) Specifically, we checked all possible features against each of the 52 concepts and wherever a feature was judged to be true of an item but was not marked in the data set, it was added. Labels that described the same feature were combined together as per Garrard et al. (2001). Finally, a few extra features were added to ensure that each item had at least one feature in each knowledge type. The statistics for the edited sets are also shown in **Table 2**. After these edits, the representational density of the Cree and McRae data set increased substantially and was now comparable to that of the other sets.

Finally, it should be noted that in both the Garrard et al. and the Cree and McRae data sets (including the edited version), the sensory/perceptual representations were denser than both the functional and the encyclopedic representations. Since this was true of both feature lists, it is very likely that it is true of people's mental modality-specific representations as well. Unfortunately, the nature of the LSA vectors did not allow us to compute a similar measure for verbal representations.

ANALYSIS AND RESULTS

CONCEPTUAL STRUCTURE WITHIN INDIVIDUAL MODALITIES

Our first goal was to assess the representational structure that each modality gives rise to. We took three distinct approaches: (1) hierarchical cluster analyses, using the Euclidean distances between pairs of items (the feature-based vectors of each concept) to build dendrograms depicting the structure in each data set; (2) correlational analyses, computing correlations between pairs of items in each data set, giving rise to correlational plots depicting the similarity structure (as opposed to distance or dissimilarity); and (3) a different computation of the similarity between concepts, this time using the cosine between pairs of vectors – for each item in each data set, we produced a list of most similar concepts. All analyses reported in this and following sections were computed in the statistical package R using standard parameter settings. The results are reported grouped by knowledge type/modality.

Sensory/perceptual representations

Figures 1 and **2** depict the plots for the four data sets within the sensory/perceptual modality. They all show a relatively detailed structure, which generally follows categorical distinctions. For example, there is a basic separation between the animals and the non-animals. Furthermore, the birds form a subcategory within the animal group. Within the non-animal group, the fruits and vegetables tend to cluster together, as do the vehicles.

Each data set comes with its own curious idiosyncrasies and interesting trends. For example, in the Rogers et al. visual set (**Figures 1A** and **2A**), the animal category forms sensible subcategories, including large land animals (*elephant*, *horse*, *camel*, and *cow*), small land animals (*rabbit*, *squirrel*, *frog*, and *mouse*), canine and felines, and reptiles. Amongst the cluster of birds, *chicken* is grouped with *owl*, and then *penguin* in the dendrogram, but the correlational plot shows that *chicken* is in fact most highly correlated with *peacock*. Within the non-animal category, in addition to the fruits and vegetables and the vehicles, there is also a large cluster of implements with coherent taxonomic substructure (*hammer* and *screwdriver*, *pliers* and *scissors*, *hairbrush* and *paintbrush*). Notably, two of the vehicles, *helicopter* and *sled*, do not group with the rest from their category in the dendrogram, but the correlational plot again shows inter-correlations among the whole class. Finally, the correlational graph reveals that the associations among the non-animals are much weaker than those among the animals.

In the Garrard et al. sensory feature dendrogram (**Figure 1B**), with the exception of the bird subcategory, the animals do not organize into subclusters as neatly as they did in the Rogers et al. set. Also, *piano* curiously clusters with the bird group instead of the non-animals (though inspection of the correlational plot proves that this item correlates most with a few of the artifacts, and even those correlations are very weak). Within the non-animals, the fruits and vegetables all group together, as do the vehicles (which form two subclusters – smaller vs. larger vehicles). While certain tools pair together as seen before (e.g., *hairbrush* and *paintbrush*), there is no coherent cluster of implements. **Figure 2B** shows that the artifacts correlate with each other, but considerably more weakly than the living things (as seen previously). In this set, the category of fruits and vegetables appears much more distinct.

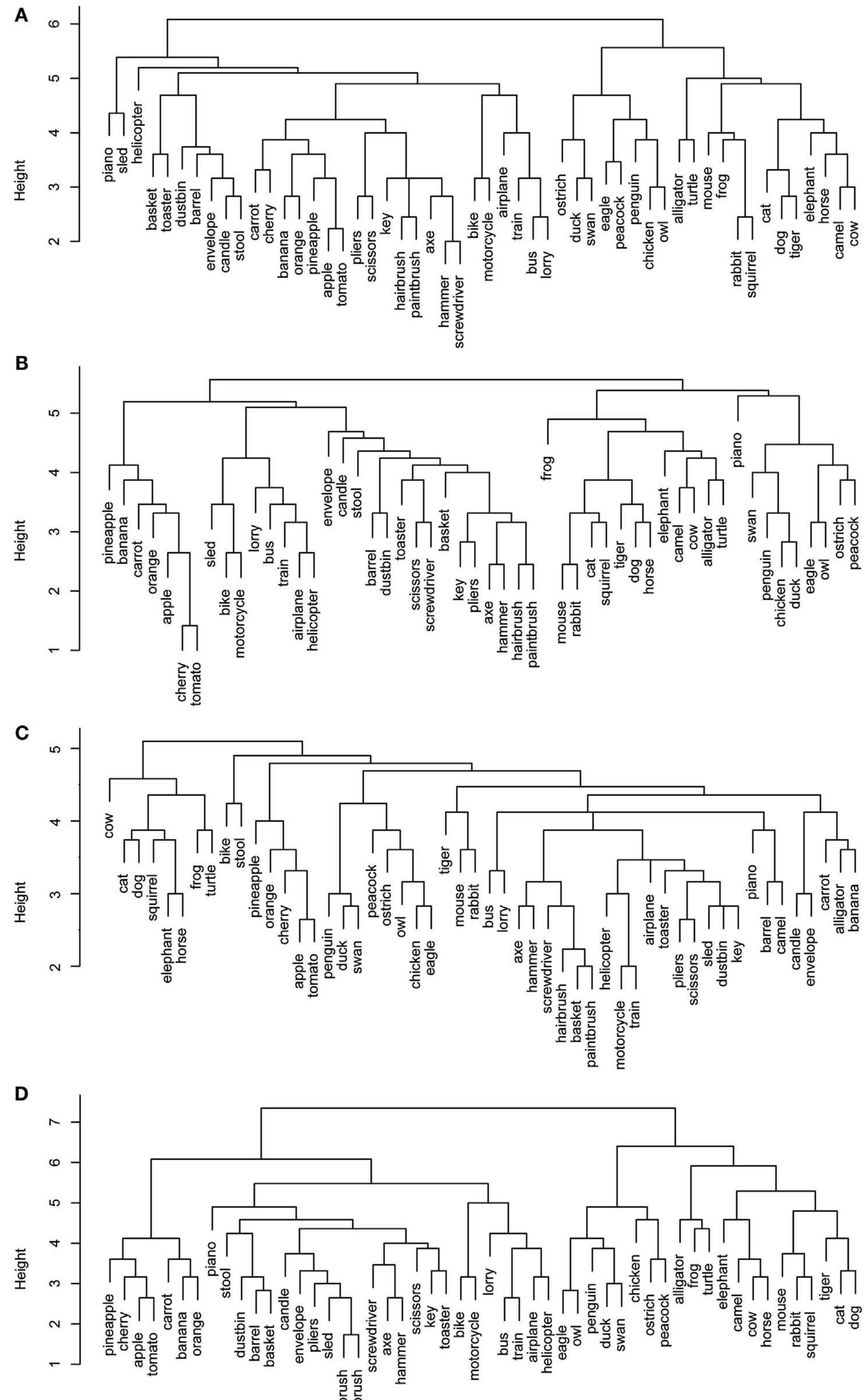


FIGURE 1 | Dendrograms from hierarchical cluster analysis for each sensory/perceptual data set: (A) Rogers et al. visual; (B) Garrard et al. sensory; (C) Cree and McRae perceptual; (D) Cree and McRae edited perceptual.

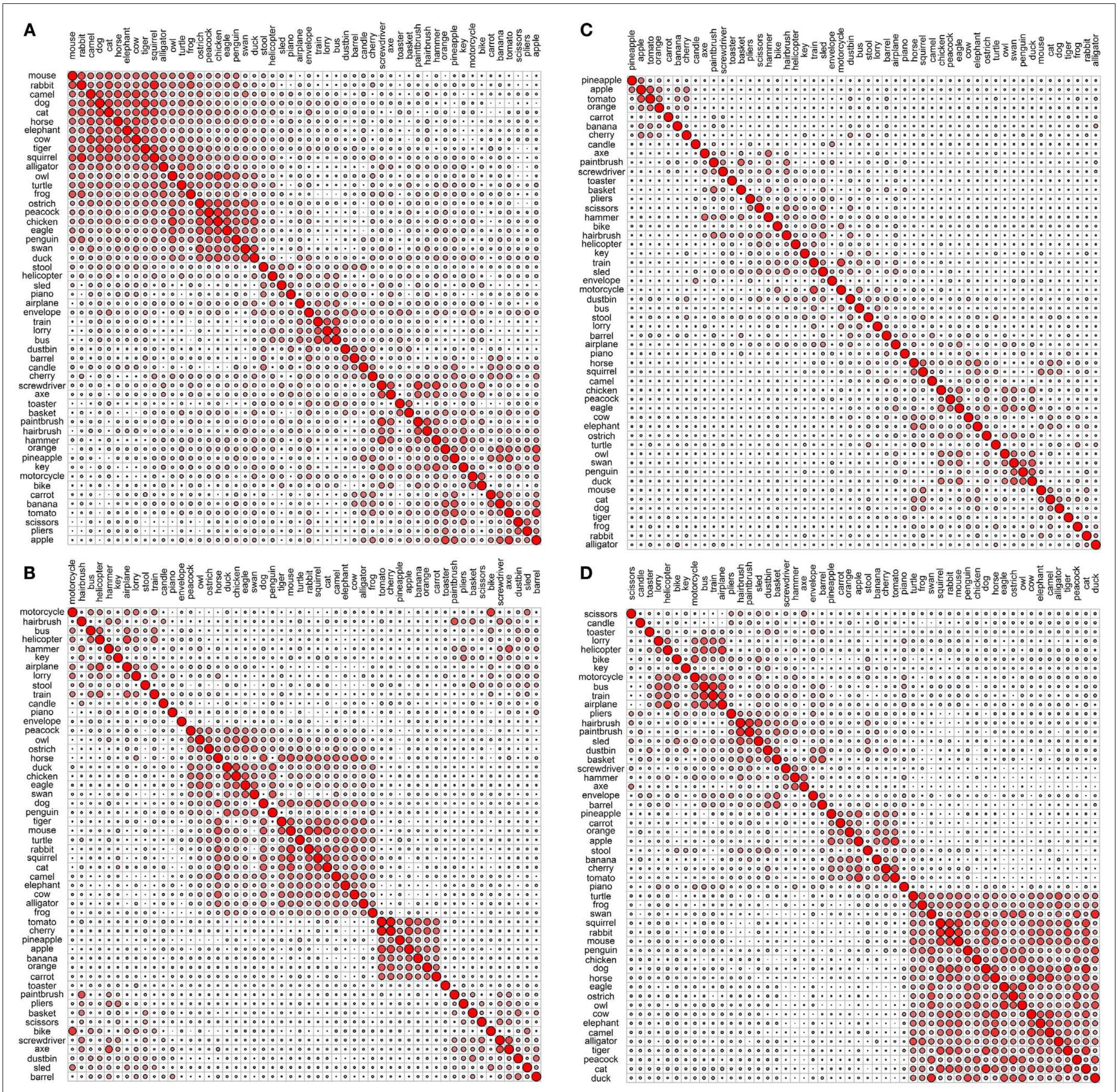


FIGURE 2 | Correlational plots for each sensory/perceptual data set: (A) Rogers et al. visual; (B) Garrard et al. sensory; (C) Cree and McRae perceptual; (D) Cree and McRae edited perceptual. The relative size of the circles represents the relative magnitude of the corresponding Pearson correlation coefficient; positive correlations are in red, negative are in blue. The ordering of the concepts along the two axes is identical and it is generated automatically to best depict clusters of concepts with high inter-correlations. As a result, the ordering may differ between graphs.

It is likely that many of the differences between the organization seen in the Rogers et al. set vs. the Garrard et al. as well as the Cree and McRae sets stem from the fact that the latter include not only static (colorless) visual information but also other perceptual features. It is also possible that drawings provide a more direct sample of true visual experience whereas feature elicitation is inevitably somewhat influenced by the demands and vocabulary-availability of speech production (see Rogers et al., 2004; Hoffman et al., 2011; Hoffman and Lambon Ralph, 2012, for further discussion of this issue).

The original Cree and McRae perceptual set exhibits the most puzzling structure. As can be seen in **Figure 1C**, the birds form their own cluster but they group with the artifacts instead of the other animals. The set of animals itself does not group entirely together – *tiger*, *mouse*, and *rabbit* form their own little cluster, which joins with the artifacts; and *camel* clusters with *barrel* and *piano*. The vehicles do not form a coherent group either – for example, *motorcycle* which was paired with *bike* in both previous sets, now pairs with *train*, while *bike* goes with *stool*. Moreover, **Figure 2C** reveals that there is very little and very weak correlational structure in this data set.

We believe that this messy and weak conceptual organization can be explained by the data set's low representational density and, relatedly, the inconsistent listing of features for some concepts but not others (even when the features are relevant). This notion is supported by comparing the structure seen in this original data set with the edited version (**Figures 1D** and **2D**). **Figure 2D** shows a clear correlational structure, especially within the animals but also within the fruits and vegetables, and more weakly within the artifacts (similar to the Rogers et al. and the Garrard et al. sets). **Figure 1D** exhibits a highly categorical organization, with the birds clustering together, and grouping with the rest of the animals, and the fruits and vegetables clustering together and attaching to the artifacts. The vehicles also cluster together (with the exception of *sled*) as do the containers (*barrel*, *basket*, and *dustbin*). Similarly to the Rogers et al. set, the animal group exhibits taxonomic substructure with four distinct categories – large land animals, small land animals, canine/felines, and amphibian/reptiles.

Overall, the perceptual representations show a relatively consistent and generally categorically organized conceptual structure – which is driven not by knowledge of categories (given that category is not coded directly in the feature vectors) but by the sheer fact that members of specific taxonomic groups tend to share perceptual attributes.

Functional representations

The first thing to note about the functional representations (**Figures 3** and **4**) is that their organization is much flatter than that of the perceptual ones. While the members of some taxonomic categories do cluster together (like fruits and vegetables, vehicles, containers, tools), members of other categories are often present in these clusters too, and there is considerably less of the categorical substructure that was observed with the perceptual representations (e.g., the birds as a subcategory of animals).

In the Garrard et al. functional set, the fruits and vegetables cluster together and are strongly intercorrelated (**Figures 3A** and **4A**). Most of the tools form a cluster as well, as do the three containers. Eight of the animals form a separate group while the other five join the birds group. Interestingly, the two water-inhabiting reptiles (*alligator* and *turtle*) pair with *penguin* (a water-inhabiting bird) and then attach to the other birds. Another curious observation is that *eagle* and *owl* (the two most prominent flyers among the birds) pair with *airplane* and *helicopter*. The other six vehicles form their own cluster as well.

The Cree and McRae functional set has the flattest dendrogram (**Figure 3B**) and a very weak correlational structure (**Figure 4B**). There are two notable groups – food, including *rabbit*, *duck*, *chicken*, *banana*, and *tomato*, which intercorrelate and also form a single cluster; transportation, including six of the eight vehicles and a pair animals (*camel* and *horse*). In the dendrogram, these items formed two separate clusters (*horse*, *bike*, and *camel* vs. *motorcycle*, *helicopter*, *bus*, *airplane*, and *train*). In addition, *orange* and *pineapple* are strongly correlated, as are *cherry* and *apple*.

The edited version of the Cree and McRae set has a much more pronounced correlational structure (**Figure 4C**). The food group has expanded to include all the fruits and vegetables as well as *rabbit*, *duck*, *chicken*, and *cow*. The artifacts all correlate together, with the transportation subcategory now including all eight vehicles and three animals (*camel*, *horse*, and *elephant*), as well as *stool* – which even though is not a means of transportation is functionally related because it is something we sit on. A group of relatively rare animals and birds also correlate together strongly.

The dendrogram shows a much richer structure than the original set as well (**Figure 3C**). As suggested by the correlational plot, the food items form a single coherent cluster as do the set of rare animals and birds (*dog* and *cat* group with these animals as well, even though they did not correlate with them). The intercorrelated transportation items formed two separate clusters (*bike*, *motorcycle*, *sled*, *horse*, and *elephant* vs. *train*, *airplane*, *bus*, and *helicopter*). In addition, the containers cluster together too, and the tools form two distinct clusters (*screwdriver*, *pliers*, and *hammer* vs. *axe*, *scissors*, and *paintbrush*).

Overall, even though we can talk about taxonomic categories within the functional representations, the concepts in these sets are clearly organized according to different principles compared to the perceptual representations – the organization here is driven by item behavior and use. For example, animals mostly group together because they do similar things (move, eat, etc.) and we do similar things with them (look at them, feed them, cook them, etc.). However, the few animals that have other uses (like food or transportation) cluster with different items, not with the animal group.

Encyclopedic representations

The encyclopedic features give rise to a conceptual structure again flatter than the one created by perceptual features, but it is also notably different from the functional organization. There tend to be smaller groups of pairs and triplets of related items. The clusters we see here generally obey the animal vs. non-animal distinction, and the fruits and vegetables tend to group together, as do the vehicles; but other than that, the organization is non-taxonomic. For example, as in the functional sets, birds are mixed with the other animals: not randomly but in interesting and predictable ways.

In the Garrard et al. encyclopedic set (**Figures 5A** and **6A**), the strongest grouping is that of the fruits and vegetables – they form a single cluster and intercorrelate strongly. In addition, the vehicles intercorrelate, but in the dendrogram they do not all group together – there is a cluster of the three large land vehicles (*train*, *bus*, *lorry*); a pairing of the smaller land vehicles (*bike* and *motorcycle*), and a pairing of the aircrafts (*airplane* and *helicopter*). The dendrogram shows a cluster of 10 household items including some tools, but the correlational plot demonstrates that, other than *axe* and *hammer*, which pair together, these items do not correlate with each other or with any other item; they are simply more or less idiosyncratic. Finally, the birds and animals mostly fall into two groups – domesticated farm animals (*horse*, *dog*, *chicken*, *cat*, and strangely *peacock*, which is also listed as domesticated), and a miscellaneous group of mostly rare animals and birds (*elephant*,

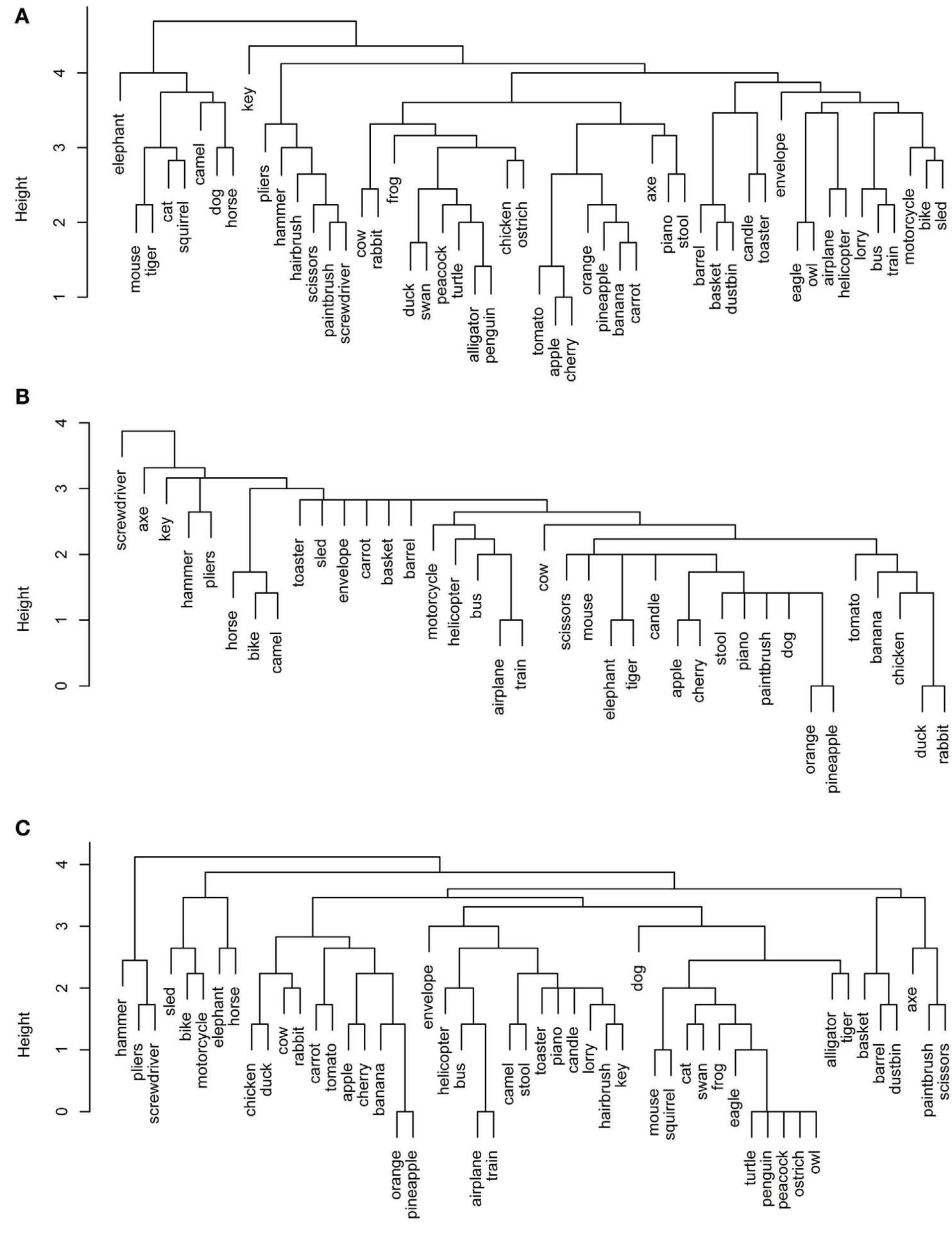


FIGURE 3 | Dendrograms from hierarchical cluster analysis for each functional data set: (A) Garrard et al. functional; (B) Cree and McRae functional; (C) Cree and McRae edited functional.

swan, *alligator*, *tiger*, *turtle*, *camel*, *ostrich*, and oddly *duck*) – these are all species one may see at the zoo.

As with the original Cree and McRae functional set, the encyclopedic set exhibits a relatively flat dendrogram and weak correlational structure (**Figures 5B** and **6B**). The fruits and vegetables are intercorrelated but they do not group together in the dendrogram. There are a few smaller clusters including a triplet of tools (*pliers*, *screwdriver*, and *hammer*), a group of reptiles/amphibians, a group of non-flying birds (*ostrich*, *peacock*, and *penguin*) and a group of five farm animals and birds. Even though the birds appear in different areas of the dendrogram, they are all intercorrelated (and correlated to *turtle* as well). There also seem to be some random pairings (e.g., *motorcycle* and *tiger*), though those do not seem to be supported by the correlational plot.

In the edited Cree and McRae encyclopedic set (**Figures 5C** and **6C**), all the fruits and vegetables are not only intercorrelated but also cluster together. There is also a more coherent group of domesticated farm animals and birds (*cat*, *dog*, *horse*, *cow*, and *chicken*). The triplet of tools that we observed in the original set is still present. The flying birds (*duck*, *swan*, *eagle*, and *owl*) cluster together, while the non-flying birds group with an interesting set of animals (for example, *penguin* goes with two other water

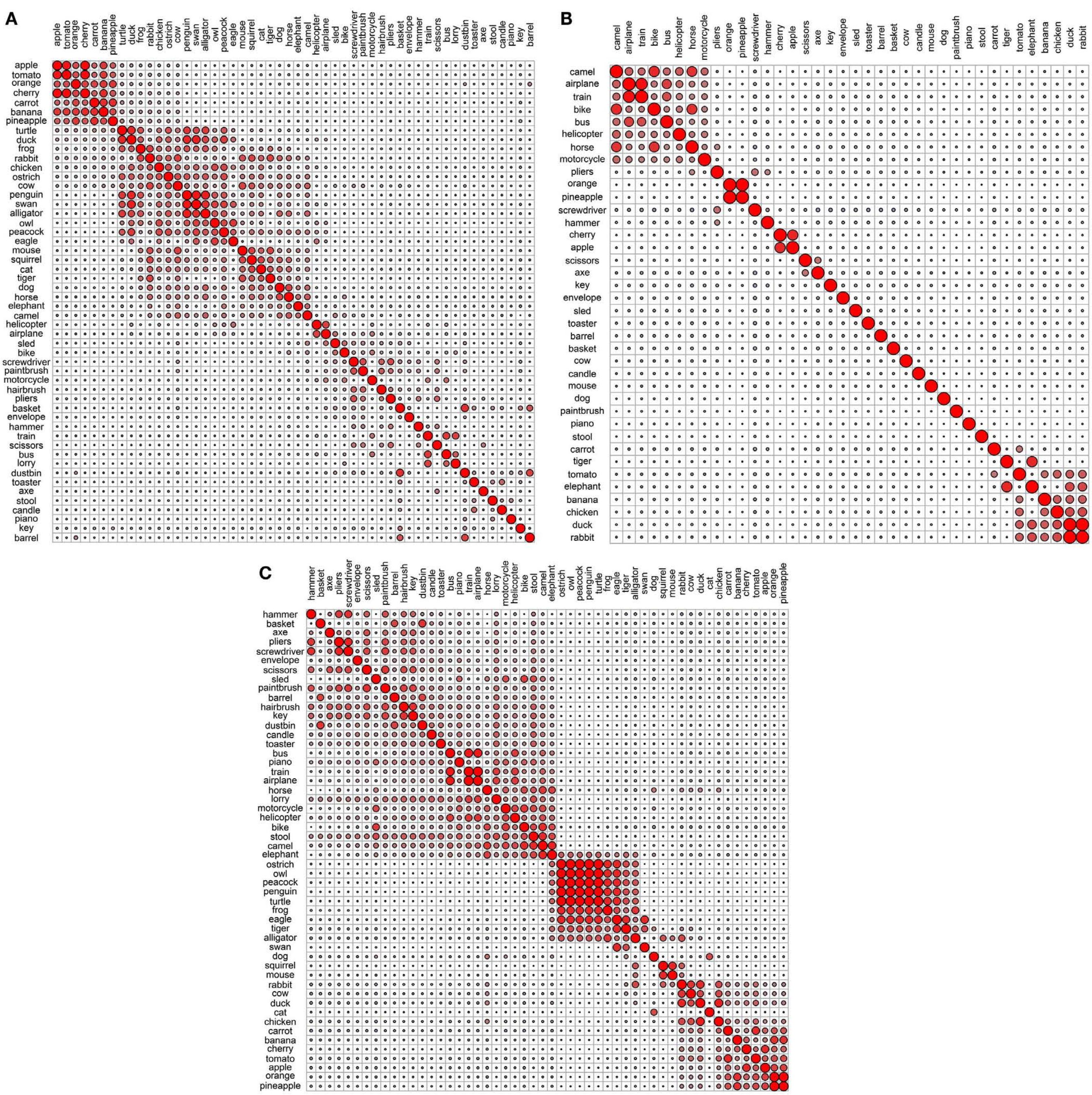


FIGURE 4 | Correlational plots for each functional data set: (A) Garrard et al. functional; (B) Cree and McRae functional; (C) Cree and McRae edited functional.

inhabitants, *alligator* and *frog*, while *ostrich* goes with *camel*). As in the original set, the birds correlate with each other, with *turtle*, and with *alligator*. It may seem strange that birds group with reptiles/amphibians but in addition to shared habitats, these two classes also share laying eggs.

Overall, the encyclopedic representations give rise to a relatively flat and more localized structure with smaller groups of items organized by relational principles such as where you normally see these items (at home, in a toolbox, on a farm, in the garden, on the road, in the sky, in the water, in the desert, and so on).

Verbal representations

The 52 verbal representations derived by Hoffman et al.'s LSA are very weakly intercorrelated (**Figure 7B**) and seem to be the set least organized by taxonomy (**Figure 7A**). There is no general differentiation between animal and non-animal items (as we saw in

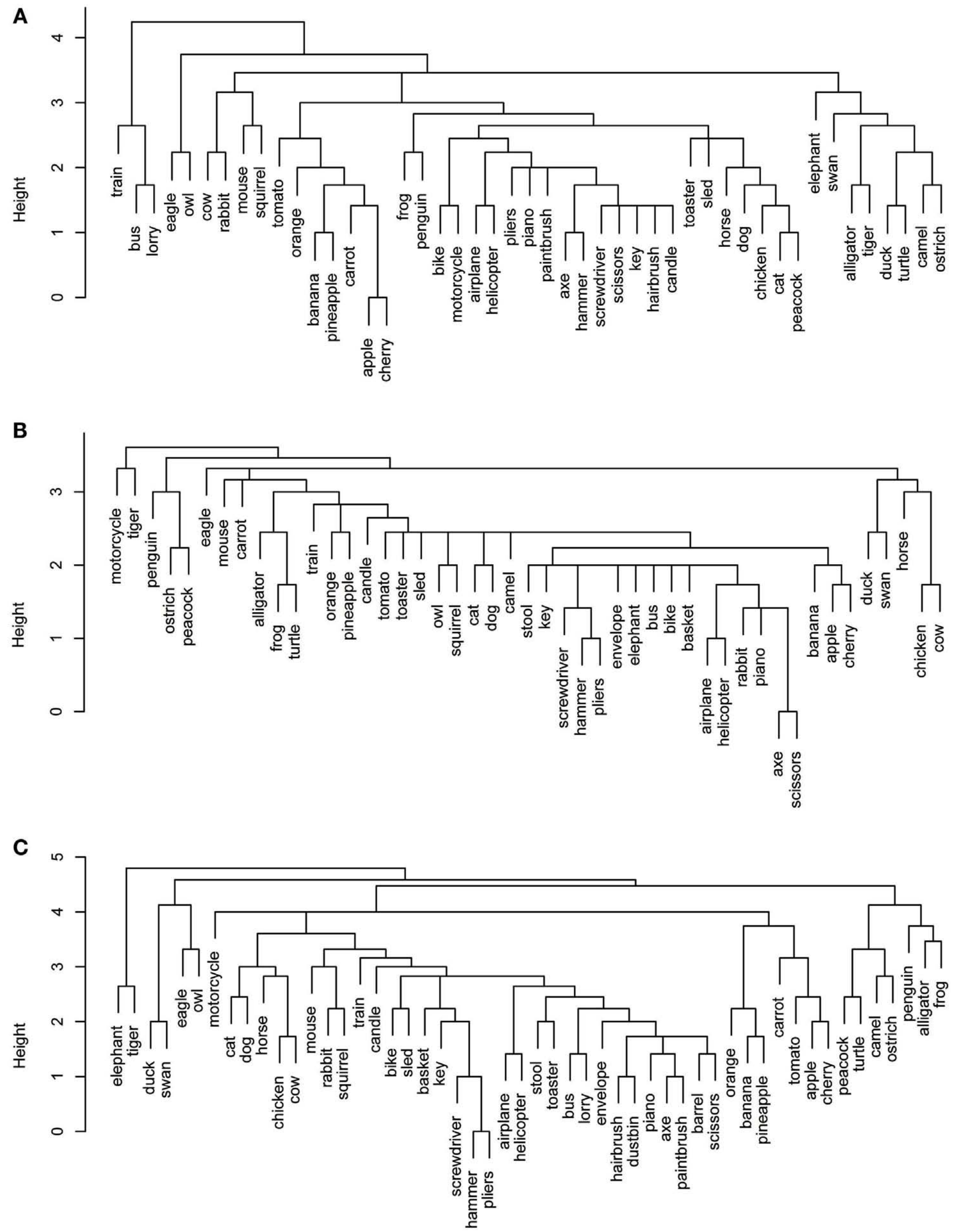


FIGURE 5 | Dendrograms from hierarchical cluster analysis for each encyclopedic data set: (A) Garrard et al. encyclopedic; (B) Cree and McRae encyclopedic; (C) Cree and McRae edited encyclopedic.

the previous sets). Nonetheless, there are some smaller categorical clusters. For example, a pair of birds (*eagle* and *owl*) and a triplet of tools (*pliers*, *screwdriver*, and *hammer*) group together and strongly intercorrelate. The four road vehicles (*bus*, *lorry*, *bike*, and *motorcycle*) also go together. *Airplane* and *helicopter* correlate, but they do not pair together in the dendrogram; *train* does not correlate with any of the other items, and interestingly, *sled* correlates most strongly with *dog* (though they do not pair together, because *dog* correlates yet more strongly with *cat*). Finally, there is a coherent cluster of food items including six of the fruits and vegetables as well as *chicken*. These items intercorrelate as well.

The dendrogram also includes a group of animals one may see on a farm (*horse*, *cow*, *dog*, *cat*, *mouse*, *rabbit*, and oddly *frog*), with the addition of *barrel* (another item not uncommon in this context). This cluster, however, is not supported by the correlational plot (other than the pairing of *dog* and *cat*, which correlate strongly). In addition to *dog* and *sled*, there are a few other interesting pairings in this set, including *elephant* and *tiger*, *hairbrush* and *scissors*, and *piano* and *key*. While one might argue that the similarity between elephant and tiger and perhaps even hairbrush and scissors is categorical in nature (perhaps due to perceptual, encyclopedic, or functional commonalities), it is crystal clear that

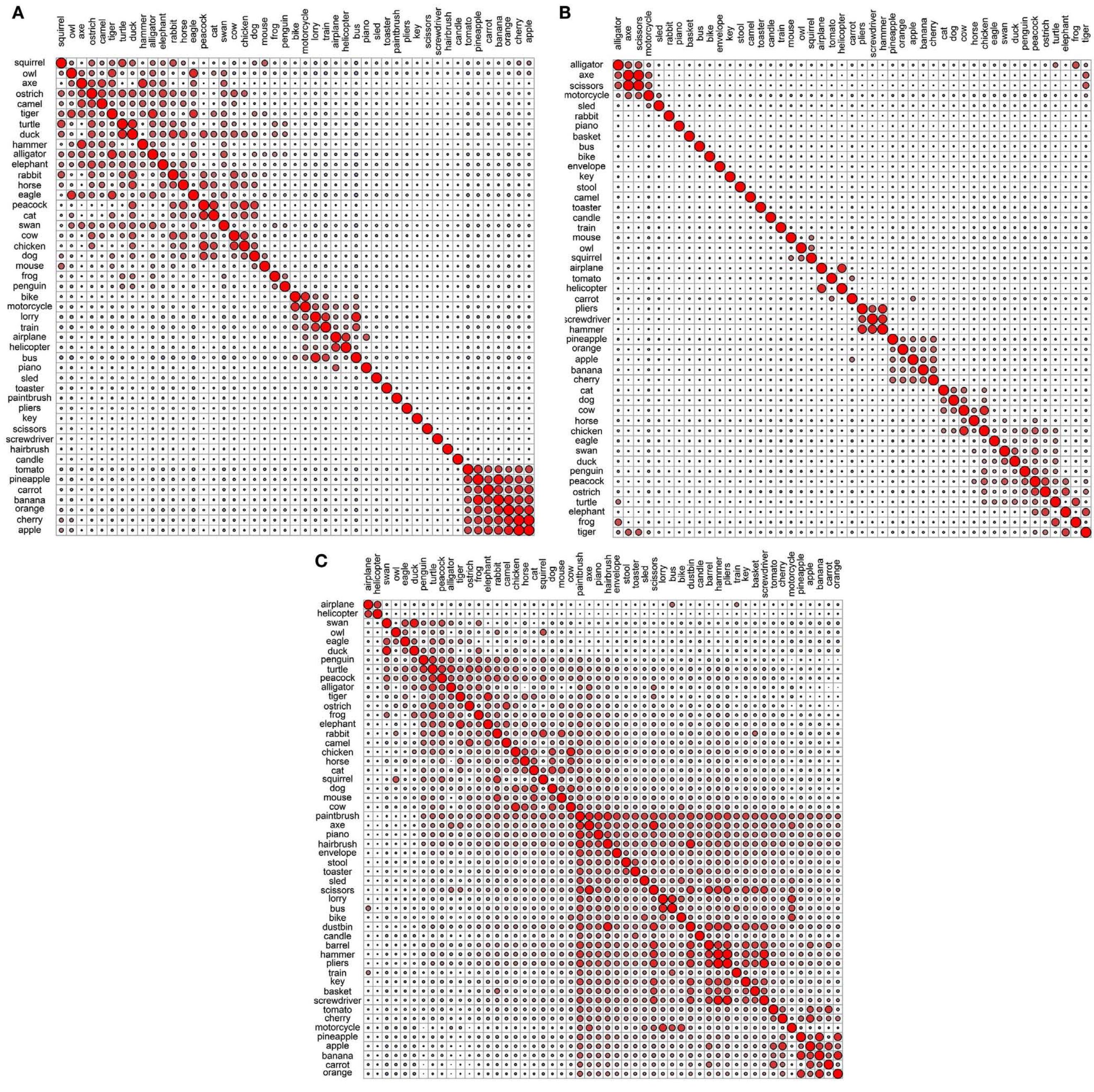


FIGURE 6 | Correlational plots for each encyclopedic data set: (A) Garrard et al. encyclopedic; (B) Cree and McRae encyclopedic; (C) Cree and McRae edited encyclopedic.

the similarity between dog and sled and piano and key is of a different – associative – nature.

Overall, the hierarchical cluster analysis and the pairwise correlations of the verbal representations illustrate that the principle governing the conceptual organization here is contextual similarity, as would be expected given the origin of these representations. Sometimes, this may be consistent with functional attributes (as items that do similar things and are used in similar ways or for similar purposes tend to appear in common verbal contexts); other times, it may be consistent with encyclopedic attributes (e.g., location, origin, behavior); and in yet other cases, it may have an associative or relational nature (e.g., *dog* and *sled*; *piano* and *key*; *hairbrush* and *scissors*).

As noted at the beginning of the section, in a final effort to assess the representational structure present in each modality-specific feature type, we investigated the most similar items for

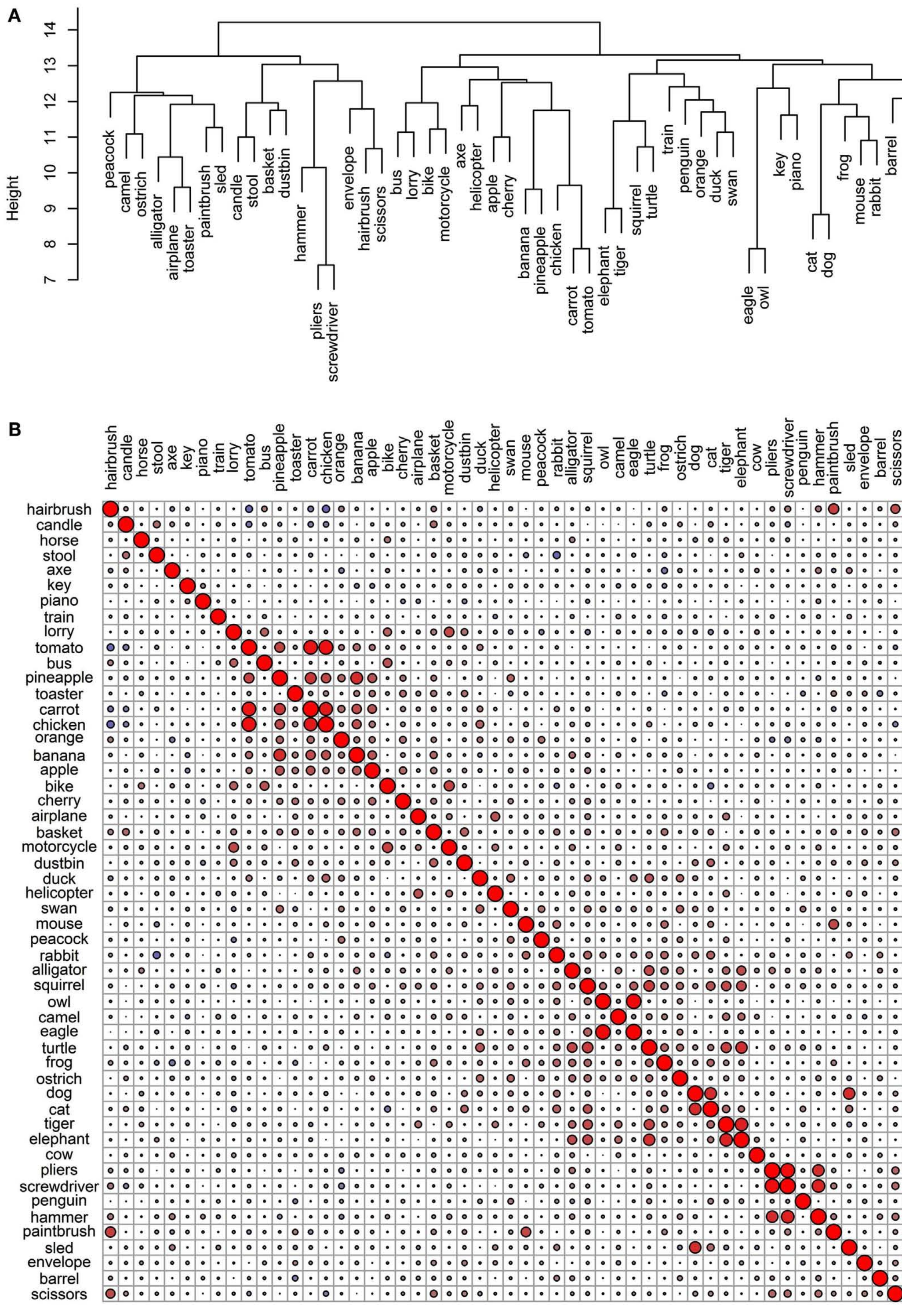


FIGURE 7 | Conceptual organization of the Hoffman et al. verbal data set: (A) hierarchical cluster analysis dendrogram; (B) correlational plot.

all concepts. To do this, we used cosine as a standard measure of similarity between two n-dimensional vectors, a and b: $cos = \frac{\sum_{i=1}^{n}(a_i b_i)}{\sqrt{\sum_{i=1}^{n}(a_i^2)\sum_{i=1}^{n}(b_i^2)}}$. The higher the cosine, the higher the similarity. This analysis progressed as follows: (1) calculate the cosine of each pair of vectors in a given set; (2) for each concept, calculate the average cosine value and its standard deviation; (3)

for each concept, generate a list of items whose cosine with the target concept is at least two standard deviations above the average for that target concept. In other words, the resulting lists indicated items that were exceptionally similar to the target concept. The full set of lists is included in Appendix. **Table 3** shows 10 concepts (two from five categories) chosen to illustrate the salient and important differences of the conceptual structure across modalities.

Chicken is perceptually most similar to other birds including *peacock*, *eagle*, and *owl*. It is functionally most similar to other birds that don't fly like *peacock* and *ostrich* (in the Garrard et al. set) as well as other animals we cook and eat like *duck* and *rabbit* (in the Cree et al. set). Encyclopedically, *chicken* is most similar to other animals domesticated farm animals (*cow*, *cat*, *dog*, *horse*) and *peacock* (which was also listed as domesticated), while verbally it is most similar to other things we eat and/or cook together with chicken, namely vegetables (*carrot* and *tomato* are the only two vegetables on our concept list). Likewise, *duck* is perceptually most similar to *swan* and *chicken*; functionally most similar to other animals that swim like *swan*, *penguin* and *turtle* (in the Garrard et al. set) as well as other animals we eat like *chicken* and *rabbit* (in the Cree et al. set). Encyclopedically, *duck* is most similar to other animals that like water (*turtle*, *swan*, *frog*); and verbally, it is also most similar to animals in the water (*turtle* and *swan*) but also other birds (*chicken* and *ostrich*).

These two examples draw attention to a discrepancy in the feature labeling between the Garrard et al. set and the Cree and McRae set. While the former included behavioral characteristics (like flying, swimming, laying eggs) in the functional knowledge type, the latter set classified those as encyclopedic features. Hence the overlap in similarity we see here.

The next pair of items are two farm animals, *cow* and *horse*. Perceptually, they are similar to each other as well as to *camel* (which is another hooved animal of similar size). Functionally, *horse* is most similar to other items that can be used to ride on like *camel*, *bike*, and *elephant*, as well as *dog* (both horses and dogs are used to pull things). *Cow*, on the other hand, is most similar to other animals we cook and eat like *rabbit*, *chicken*, and *duck* (in the Cree and McRae set); in the Garrard et al. set, it appears as most similar to *mouse* and *rabbit*, but this again is an artifact of the varying classification [in addition to being edible – which is something *cow* shares with *rabbit* – the three animals share the ability to breed, chew, eat, walk, and run, which are all features that would be classified as perceptual or encyclopedic in nature according to Cree and McRae (2003)]. Encyclopedically, cow and horse are similar to each other as well as other domesticated farm animals (*chicken*, *dog*, and *cat*). Verbally, the two differ as well – *horse* tends to appear in verbal contexts similar to *bike* (probably due to their shared function), while *cow* tends to appear in verbal contexts shared with *elephant* (this latter finding was somewhat surprising).

The next pair of examples comes from the fruit-and-vegetable category. Perceptually, *carrot* is most similar to same category members with common shape and/or color (*banana* and *orange*) as well as other items with an elongated shape. Likewise, *apple* is most similar to *tomato*, *orange*, *cherry*, and *pineapple*. Functionally, they also relate most highly to same category members, notably *carrot* is most similar to the other vegetable in the set (*tomato*) as well as some fruit we use to make juice (*orange* and *pineapple*), while *apple* is most similar to *cherry* (both being fruits we grow on trees in our gardens and pick, and use to make pie!) Encyclopedically, both *carrot* and *apple* have most in common with other things we commonly grow in our gardens (each other, as well as *tomato*); in addition, *apple* is similar to other fruits that grow on trees (*cherry*, *banana*, *orange*, and *pineapple*). Verbally, both items share contexts with their own sets of food items (*tomato*, *chicken*, and *pineapple* for *carrot*; *pineapple*, *carrot*, and *banana* for *apple*).

Moving to the artifacts, *bus* and *bike* are most similar to other vehicles in all four feature types but in subtly different ways. Perceptually, *bus* shares the most with other large vehicles (*lorry*, *train*, and *airplane*), while *bike* is most similar to the smaller vehicles (*motorcycle* and *sled*). Functionally, *bus* is most similar to other vehicles used for public transportation (*train* and *airplane*), while *bike* is most similar to other things we can sit on and ride – not only vehicles (*motorcycle* and *sled*) but also animals (*camel* and *horse*). Encyclopedically, both *bus* and *bike* have most in common with other vehicles seen on the road (*lorry* and *motorcycle*), which also seems to be the most commonly shared verbal context (as *bus* is verbally most similar to *bike* and *lorry*, and *bike* is verbally most similar to *motorcycle*, *bus*, and *lorry*). *Bike* and *motorcycle* have a very high verbal similarity also due to the fact that they are often used synonymously.

The final example is a pair of tools, which – like the vehicles – are most similar to other items from their category in all four knowledge types. Perceptually, *screwdriver* has most in common with *hairbrush*, *hammer*, and *axe* (they all have a single handle and, with the exception of axe, similar size); likewise, *pliers* has most in common with *scissors*. Functionally, both items are very similar to implements used in a handheld manner for handiwork (*screwdriver*, *pliers*, *hammer*, and *paintbrush*); in addition, *pliers* are like *scissors* in that they can cut. Encyclopedically as well as verbally, *screwdriver* and *pliers* are also most similar to each other and *hammer* (which is another item commonly found in a toolbox), but not any of the other implements that appeared in the perceptual and functional sets.

In summary, this analysis supports the results obtained in the hierarchical cluster analyses and the correlations – the conceptual representations within the investigated four knowledge types organize in unique and sensible ways – within the perceptual modality, conceptual structure is governed by perceptual similarity (most prominently shape, size, color, and parts); within the functional modality, the structure is directed by similarity in use and interaction; within the encyclopedic modality, it obeys commonality in location, habitat, and/or behavior; within the verbal modality, it is associative or relational (similarity within the verbal domain may be functional or encyclopedic in nature but need not be).

COMPARING CONCEPTUAL STRUCTURE BETWEEN MODALITIES

The second step in our investigation was to determine how conceptual structure compares across modalities. To do this, we took the distance matrices for each set – that is, the set of Euclidean distances between each pair of concepts within a given data set (these same matrices were used in the hierarchical cluster analyses discussed above) – and computed the pairwise matrix correlations

Table 3 | Lists of items most similar to a set of 10 concepts in each of the 11 data sets.

(A)

Concept	Rogers et al. visual	Garrard et al. sensory	Cree and McRae perceptual	Cree and McRae edited perceptual	Hoffman et al. verbal
Chicken	Peacock, owl, eagle, ostrich	Duck, eagle, owl	Eagle, duck, owl, swan	Peacock	Tomato, carrot
Duck	Swan, chicken, peacock, ostrich	Chicken, penguin	Swan, penguin, eagle, chicken, owl	Swan	Turtle, swan, chicken, ostrich
Cow	Camel, horse, dog	Horse, camel	Horse, elephant	Horse, camel	Elephant
Horse	Cow, camel	Mouse, dog	Elephant, squirrel, cat	Cow, camel	Bike
Carrot	Banana, pineapple, candle, cherry	Apple, cherry, tomato, orange, banana	Paintbrush, screwdriver, banana, alligator, orange, tiger	Orange, banana, cherry, apple, pineapple	Tomato, chicken, pineapple
Apple	Tomato, orange, banana, pineapple	Cherry, tomato, carrot, orange, banana, pineapple	Tomato, cherry, pineapple, orange	Tomato, cherry, pineapple, banana	Pineapple, carrot, banana
Bus	Lorry, train, airplane	Helicopter, airplane, train, motorcycle	Motorcycle, lorry, train	Train, airplane, helicopter, lorry, motorcycle	Bike, lorry
Bike	Motorcycle, screwdriver	Motorcycle, sled, bus	Motorcycle, train, sled	Motorcycle, sled	Motorcycle, bus, lorry
Screwdriver	Paintbrush, hammer, axe	Axe, hairbrush, scissors	Hairbrush, hammer, paintbrush, axe	Hammer, paintbrush, axe	Pliers, hammer
Pliers	Scissors, orange, apple	Key, hammer, axe	Scissors	Hairbrush, sled	Screwdriver, hammer

(B)

Concept	Garrard et al. functional	Cree and McRae functional	Cree and McRae edited functional	Garrard et al. encyclopedic	Cree and McRae encyclopedic	Cree and McRae edited encyclopedic
Chicken	Owl, peacock, ostrich	Duck, rabbit, banana, tomato	Duck, cow, rabbit	Peacock, cow, cat	Cow, peacock	Cow, dog, horse
Duck	Swan, penguin, turtle, peacock	Chicken, elephant, banana, tomato	Chicken, rabbit	Turtle, rabbit	Swan, turtle, peacock	Swan, frog, turtle
Cow	Mouse, rabbit	NONE	Rabbit, chicken, duck	Chicken, rabbit, horse	Chicken, cat, dog	Chicken, dog, horse
Horse	Dog, camel	Bike, camel	Camel, elephant, bike	Duck, peacock, cow	Cow, peacock, chicken, swan	Cow, chicken, cat, dog
Carrot	Banana, pineapple, orange	Tomato	Tomato, orange, pineapple	Apple, banana, cherry, pineapple, orange, tomato	Apple, tomato	Apple, tomato
Apple	Cherry, tomato, banana	Cherry	Cherry, orange, pineapple, tomato	Orange, banana, carrot, pineapple, tomato	Cherry, banana, orange, pineapple, carrot	Cherry, carrot, banana, tomato
Bus	Train, lorry, motorcycle	Airplane, train	Airplane, train, helicopter	Lorry, train	NONE	Lorry, motorcycle, paintbrush
Bike	Sled, motorcycle	Camel, horse	Camel, motorcycle, sled, horse	Motorcycle, lorry, train	NONE	Motorcycle, lorry, paintbrush, sled
Screwdriver	Paintbrush, pliers, hairbrush	Pliers, hammer	Pliers, hammer, paintbrush, scissors	NONE	Hammer, pliers	Hammer, pliers
Pliers	Screwdriver, scissors, paintbrush, hammer	Screwdriver, hammer, horse	Hammer, screwdriver, paintbrush, scissors	NONE	Hammer, screwdriver	Screwdriver, barrel, dustbin, scissors

(a) Similarity lists for the sensory/perceptual and the verbal data sets. (b) Similarity lists for the functional and encyclopedic data sets.

Table 4 | Pairwise correlations between the distance matrices within each data set as well as compared to simple taxonomic structure and random structure.

(A)

	Garrard et al. sensory	Cree and McRae perceptual	Garrard et al. functional	Cree and McRae functional ($n = 38$)	Garrard et al. encyclopedic	Cree and McRae encyclopedic ($n = 47$)	Hoffman et al. verbal	Taxonomic structure	Random structure
Rogers et al. visual	0.55***	0.35*	0.27	−0.13	0.40**	0.26	0.06	0.40**	−0.03
Garrard et al. Sensory		0.52***	0.37*	−0.15	0.29	0.15	−0.05	0.60***	0.02
Cree and McRae perceptual			0.19	−0.32*	0.24	0.19	0.03	0.32*	0.01
Garrard et al. functional				0.11	0.14	0.01	0.30*	0.31*	0.01
Cree and McRae functional ($n = 38$)					−0.13	−0.16	−0.03	−0.04	−0.01
Garrard et al. encyclopedic						0.27	0.12	0.23	−0.06
Cree and McRae encyclopedic ($n = 47$)							0.07	0.1	−0.02
Hoffman et al. verbal								−0.03	−0.04

(B)

	Garrard et al. sensory	Cree and McRae edited perceptual	Garrard et al. functional	Cree and McRae edited functional	Garrard et al. encyclopedic	Cree and McRae edited encyclopedic	Hoffman et al. verbal	Taxonomic structure	Random structure
Rogers et al. visual	0.55***	0.72***	0.27	0.08	0.40**	0.43**	0.06	0.40**	−0.03
Garrard et al. Sensory		0.65***	0.37*	0.10	0.29	0.42**	−0.05	0.60***	0.02
Cree and McRae edited perceptual			0.42**	0.13	0.36*	0.57***	0.22	0.46**	−0.01
Garrard et al. functional				0.33*	0.14	0.14	0.30*	0.31*	0.01
Cree and McRae edited functional					0.02	−0.03	0.13	0.28*	0.01
Garrard et al. encyclopedic						0.43**	0.12	0.23	−0.06
Cree and McRae edited encyclopedic							0.10	0.30*	−0.01
Hoffman et al. verbal								−0.03	−0.04

*(a) With the original Cree and McRae feature lists; (b) with the edited Cree and McRae feature lists. ***$p < 0.0005$, **$p < 0.005$, *$p < 0.05$.*

for all data sets. The analysis was done twice – first with the original Cree and McRae feature list and then with the edited version. The results are presented in **Table 4**.

The most obvious finding is that the three sensory/perceptual sets intercorrelate. The correlations between the Cree and McRae set and the other two are greatly improved in the edited version, especially with the Rogers et al. set where the correlation doubles. There is also a notable trend for the conceptual structure of this type of representations to correlate with that of the encyclopedic knowledge type (**Figure 4B**), but much less so with the functional knowledge type. In fact, the original Cree and McRae perceptual and functional sets correlated negatively. The Garrard et al. functional set appears to organize more consistently with the one in the sensory modality, which probably has to do with the specific classification of features, whereby statements that had to do with behavior (e.g., lays eggs, swims, dives, runs, flies, jumps) were included in the functional feature list as opposed to the perceptual (motion) or encyclopedic feature lists, as was done by Cree and McRae. Finally, the conceptual structure present in the sensory/perceptual modality did not correlate at all with that in the verbal modality.

The conceptual networks in the functional and encyclopedic modality are considerably more idiosyncratic than that in the perceptual modality, as evident by the fewer significant correlations involving these sets (**Table 4a**). Using the edited Cree and McRae sets, we found that the concepts in the two functional sets organize in similar ways (in fact, this was the only significant correlation involving the Cree and McRae functional structure). The same is true for the two encyclopedic sets. These two types of knowledge did not correlate with each other, though the functional modality had a tendency to relate to the structure in the verbal modality. Finally, as noted above, the structure within the encyclopedic modality also correlated with that in the perceptual modality.

The conceptual organization within the verbal modality appears to be most unique. The only significant correlation (and still pretty small in magnitude) was with the functional sets (Garrard et al. and the edited Cree and McRae). This finding is consistent with the observations we made based on the dendrograms, correlational plots, and item similarity lists.

In summary, by statistically comparing how concepts relate to each other in each data set, we established that the various sets, which came from different sources, are mostly consistent with each other (with the slight exception of the Garrard et al. functional list). Furthermore, we found how conceptual structure relates across modalities. The organization within the perceptual modality is somewhat similar to that in the encyclopedic modality, and both of those are different from the functional and verbal modality. There is also some similarity between the functional and verbal knowledge types, but generally, those appear to be organized in idiosyncratic ways.

COMPARING CONCEPTUAL STRUCTURE TO SIMPLE CATEGORICAL STRUCTURE

The final goal of this investigation was to determine the degree to which the representational networks within each modality were taxonomically organized (i.e., organized according to category). To do this, we generated a simple "reference" categorical structure (**Figure 8**) where the animals, the artifacts, and the fruits and vegetables all form distinct clusters; the birds are a subcategory of the animals, and the vehicles are a subcategory of the artifacts. We computed the pairwise correlations of the distance matrix in this simple categorical structure and the distance matrices in the 11 data sets. As can be seen in **Table 4**, the sensory/perceptual modality is the only one that consistently correlated with the taxonomic structure, with the Garrard et al. sensory set having the highest correlation.

We then assessed whether the correlations between the sets and the taxonomic reference structure were significantly different from each other, taking into account the between-set correlations. The results are shown on **Table 5** and they confirmed that taxonomic organization is most prominent in the Garrard et al. sensory set. The other correlations did not systematically differ from each other, with the exception of the verbal modality, where the lack of categorical organization was significantly different from all sets which had notable categorical structure (the four sensory/perceptual sets, the Garrard et al. functional set, and the Cree and McRae edited functional and encyclopedic sets).

COMPARING CONCEPTUAL STRUCTURE TO RANDOM STRUCTURE

In the modalities where concepts are not organized taxonomically, is the structure random? To assess this, we compared the distance matrices to 1000 random permutations of the distance matrix from the taxonomic structure. The average correlation values are presented in the last column of **Figure 4** and none of them are significant, confirming our previous observation that even though concepts do not group according to category in all modalities, there is coherent structure in each case.

DISCUSSION

This study employed a number of different data sets and analytical methods in order to assess the structure of information arising in four modalities of knowledge: perceptual, functional, encyclopedic, and verbal. We had three distinct goals in mind: (1) to establish the organization in each modality; (2) to compare the structure between the various modalities; and (3) to assess the degree to which each structure is taxonomic or random.

In summary, our results showed that there is abundant structure in each of the four modalities we investigated (none of it is random) but the organization differs across modalities. The visual/perceptual domain is the most hierarchically organized and closest to classic taxonomic structure. Items group into categories and subcategories based on their prominent sensory characteristics (most importantly, shape, size, color, and parts). The organization in this modality is measurably different from the one in the functional modality, where concepts organize according to experience of interaction and use. Generally, this does *not* correlate with perceptual experience, though *occasionally* it might (e.g., in the case of some tools, which are visually as well as functionally similar). Encyclopedic knowledge gives rise to yet another conceptual organization, governed by experience of shared location or behavior. This type of structure appears to correlate with the organization within the perceptual domain (at least for this set of items) but not the functional domain. Finally, the verbal modality has the most unique structure, not at all categorical but also not random. It centers on associative or relational knowledge. It weakly resembles functional organization but notably deviates from perceptual and/or encyclopedic organization.

The findings from the current and previous studies all underline the fact that concepts are formed from a rich multimodal (verbal and non-verbal) set of experiences, where concepts relate to each other and organize in distinct ways. This, in turn, raises the question of how all these modality-specific experiences are fused together into coherent cross-modal conceptual knowledge which is capable of appropriate generalization across exemplars.

Some contemporary and classical theories postulate that the semantic system is simply this distributed network of modality-specific representations, all connected to one another (e.g., Eggert, 1977; Martin, 2007). Each modality-specific element within this distributed network would be able to code the local statistics (information structure) arising in that modality. There is a clear danger, however, that this system alone would lack knowledge of feature co-occurrence statistics across modalities (e.g., things that have beaks usually can fly, they lay eggs, nest in the trees, and sing songs). Without knowledge of the cross-modal coherent covariation of information, the semantic system would be unable to pull together the correct subset of cross-modal features for each concept and to generalize this information appropriately across concepts (Smith and Medin, 1981; Wittgenstein, 2001; Rogers and McClelland, 2004; Lambon Ralph et al., 2010). Extracting this kind of statistics is more than a simple linear summation of the individual modalities or learning pairwise correlations (see Rogers and McClelland, 2004, for a computational demonstration; and Lambon Ralph et al., 2010, for further discussion).

In keeping with these observations, a recent investigation employed a graph-theoretic approach to look at the relative

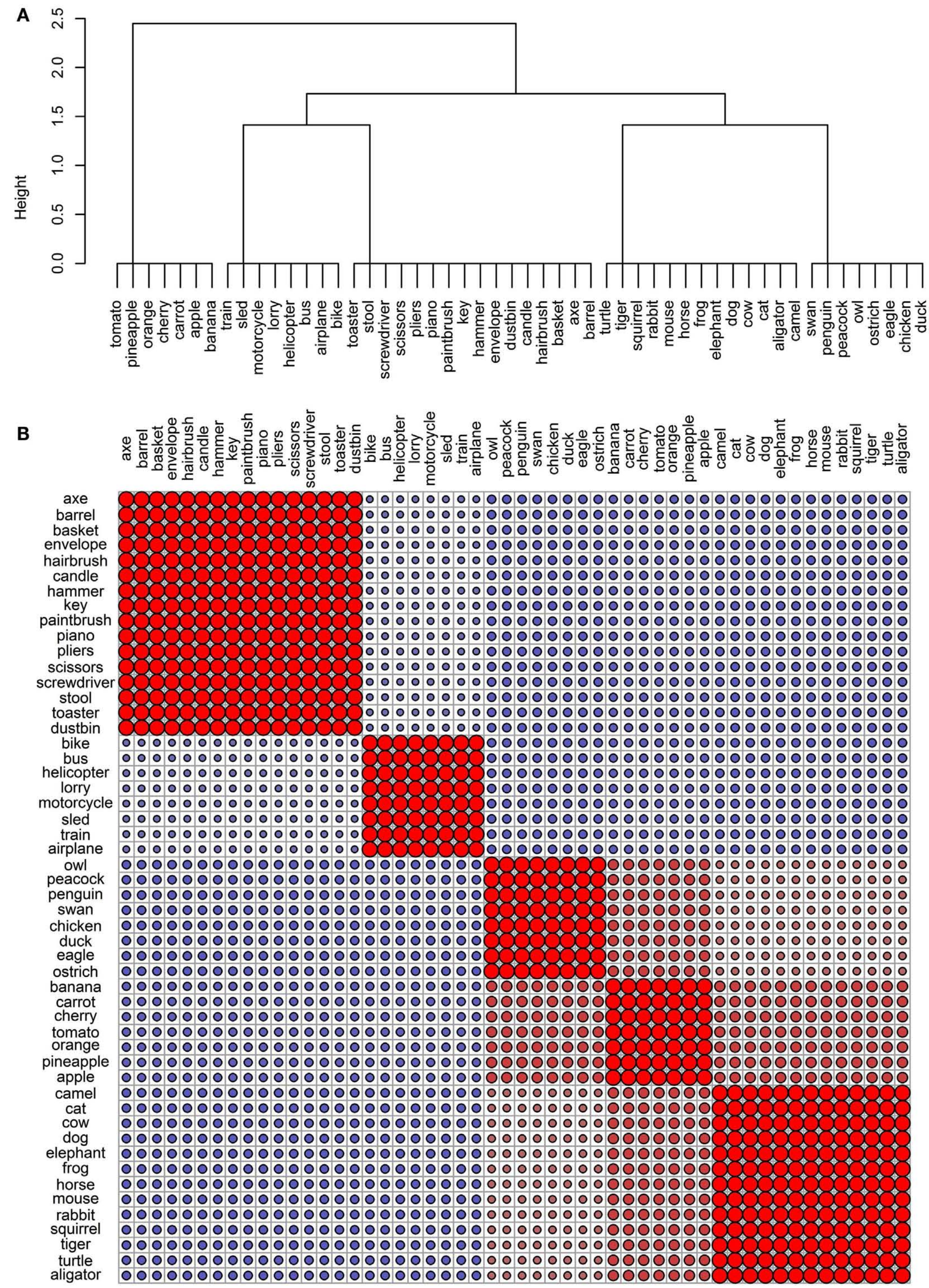


FIGURE 8 | Conceptual organization of a simple taxonomic structure: (A) hierarchical cluster analysis dendrogram; (B) correlational plot.

contribution of perceptual and functional knowledge (based on Cree and McRae's feature lists) to the conceptual organization of 130 common nouns, all acquired by 30 months of age (Hills et al., 2009). Hills and colleagues constructed three types of conceptual network – using the full set of features, using only the perceptual features, and using only the functional features. By calculating the average clustering coefficient for nodes in each network (that is, their tendency to share features with neighboring nodes), they

Table 5 | Correlations between the distance matrices in each data set and the correlation matrix in a simple taxonomic structure.

(A)

Data set	*r* Value	Garrard et al. sensory	Cree and McRae perceptual	Garrard et al. functional	Cree and McRae functional ($n=38$)	Garrard et al. encyclopedic	Cree and McRae encyclopedic ($n=47$)	Hoffman et al. verbal
		0.60***	**0.32***	**0.31***	**−0.04**	**0.23**	**0.1**	**−0.03**
Rogers et al. visual	0.40**	✓			✓		✓	✓
Garrard et al. sensory	0.60***		✓	✓	✓	✓	✓	✓
Cree and McRae perceptual	0.32*							✓
Garrard et al. functional	0.31*							✓
Cree and McRae functional ($n=38$)	−0.04							
Garrard et al. encyclopedic	0.23							
Cree and McRae encyclopedic ($n=47$)	0.1							

(B)

Data set	*r* Value	Garrard et al. sensory	Cree and McRae edited perceptual	Garrard et al. functional	Cree and McRae edited functional	Garrard et al. encyclopedic	Cree and McRae edited encyclopedic	Hoffman et al. verbal
		0.60***	**0.46****	**0.31***	**0.28***	**0.23**	**0.30***	**−0.03**
Rogers et al. visual	0.40**	✓						✓
Garrard et al. sensory	0.60***			✓	✓	✓	✓	✓
Cree and McRae edited perceptual	0.46**							✓
Garrard et al. functional	0.31*							✓
Cree and McRae edited functional	0.28*							✓
Garrard et al. encyclopedic	0.23							
Cree and McRae edited encyclopedic	0.30*							✓

Ticks mark values that are significantly different from each other ($p < 0.05$), (a) With the original Cree and McRae feature lists; (b) with the edited Cree and McRae feature lists.

evaluated the existence of structure in the networks (groups and subgroups of nodes). Their results closely resembled what we found here – there was significant structure (compared to random) in all three networks; the perceptual network was much denser than the functional network and clusters in the latter were smaller in size. Like ours, their results also indicated that these two types of features contribute differently to category organization. Furthermore, Hills et al. showed that despite the differences in the organization of these two modalities, the two types of features have a high degree of correspondence (or coherent covariation), creating conceptual structure above and beyond the structure existing in each one modality.

Recently, a number of investigations have focused on how verbal knowledge can be combined with feature type knowledge. For example, Steyvers (2010) presented a probabilistic model in which a text-based data-driven approach to extracting semantic information is augmented with knowledge of perceptual, functional, and encyclopedic features for a set of 287 animate and inanimate concepts. The results showed that the addition of feature information improved the model's ability to generalize. Similarly, Durda et al. (2009) reported a neural network model trained with 445 concepts to map from textual co-occurrence vectors (similar to the verbal representation analyzed in our study) to feature representations based on the Cree and McRae norms. Like Steyver's model, this model also exhibited a notable ability to generalize to novel concepts (i.e., ones it had not been trained on).

In an impressive computational modeling study, Andrews et al. (2009) adopted a probabilistic approach to extract semantics from a data set including both verbal and non-verbal (i.e., feature) information. The authors referred to these two types of information as distributional and experiential, respectively, to emphasize the point that language-based knowledge is qualitatively different from sensory-functional type of knowledge. Their distributional data set included about 8,000 short texts from the British National Corpus (each 150–250 words long), while their experiential data set included feature norms for 456 concepts. The model was trained using either set alone, the two sets in conjunction, or the two sets independently. In line with our findings, Andrews and colleagues observed that the semantic structure learned from the two types of information is markedly distinct, and further distinct from (and not as rich as) the structure arising when the two types of information are combined. Notably, they found that the structure in the two models trained with a single data set correlated higher with the structure in the model trained with the two sets independently than the one trained with the two sets jointly, suggesting that the conceptual organization arising under simultaneous exposure to multiple information sources is unique and different from the one arising from a single source or a linear combination of the multiple sources.

To assess performance, the model's learned semantic similarity between concepts was compared to human behavioral data including lexical substitution errors, word-association norms, lexical priming, and semantic errors in picture naming. The results showed that the combined model was most similar to the behavioral data. The authors discussed the importance of both verbal and non-verbal information in the acquisition of semantic knowledge, and emphasized the point that cross-modal semantic representations rely on exposure to the statistical structure (what we earlier called coherent covariation) both within and *between* modalities.

All these findings and observations imply that additional computational machinery is required to fuse modality-specific information together to form coherent concepts. One possibility is provided by the hub-and-spoke account (e.g., Rogers et al., 2004; Patterson et al., 2007; Lambon Ralph et al., 2010; Pobric et al., 2010) which inherits the basic premise that the multiple verbal and non-verbal modalities provide the raw ingredients for the formation of concepts (they are the "spokes" of the semantic system) but there is an additional component (the "hub" of the system) that mediates between the various modalities. The representations learned in the hub are based on *complex non-linear mappings* among the modality-specific representations in the spokes (Wittgenstein, 2001; Lambon Ralph et al., 2010). Just as modality-specific knowledge pools have been localized to distinct areas in the brain, so has the proposed transmodal representational hub – with the anterior inferolateral temporal area being one crucial region. The clearest neuropsychological example of this comes from investigations of semantic dementia, where the patients' maximal damage in this region leads to multimodal yet selective semantic impairment (Warrington, 1975; Snowden et al., 1989; Hodges et al., 1992; Bozeat et al., 2000). Convergent evidence for the importance of this region in semantic representation has been provided by functional imaging and repetitive transcranial magnetic stimulation studies with neurologically intact individuals (e.g., Vandenberghe et al., 1996; Marinkovic et al., 2003; Pobric et al., 2007; Binney et al., 2010; Visser and Lambon Ralph, 2011; Peelen and Caramazza, 2012; Visser et al., 2012). In addition, the location and connectivity of the transmodal hub has recently been established using diffusion-weighted tractography in neurologically intact participants (Binney et al., 2012) and its breakdown in semantic dementia (Acosta-Cabronero et al., 2011).

Two aspects of the hub-and-spoke theory are crucial. The first is that the transmodal hub provides the neural machinery to compute the non-linear mappings required for the formation of coherent concepts and their generalization on the basis of semantic rather than superficial (modality-specific) similarities. Indeed, recent targeted investigations of semantic dementia have shown that, in both verbal and non-verbal domains, the patients lose the coherence of these concepts and thus exhibit over- and under-generalization errors (Lambon Ralph and Patterson, 2008; Lambon Ralph et al., 2010; Mayberry et al., 2011). The second crucial aspect of this theory is that semantic representations require the combination of transmodal (hub) *and* modality-specific information sources. It is not, therefore, an issue of debating whether semantic representations are underpinned by hub OR spokes but rather how these work together to form coherent concepts. The importance of both elements is indicated in recent functional neuroimaging studies (e.g., Visser and Lambon Ralph, 2011; Peelen and Caramazza, 2012; Visser et al., 2012) and confirmed by probing hub-and-spoke regions in the same participants using rTMS (e.g., Pobric et al., 2010). In such circumstances, transient suppression of the transmodal ATL hub generates a pan-category semantic impairment whereas stimulation of the dorsal aspects of the inferior parietal lobule generates a category-specific impairment for manmade objects that relates directly to the suppression of praxis information that is coded in this region.

A recent behavioral paper focused around the issue of how concept-relevant information from different modalities is combined into coherent and useful cross-modal semantic representations (McNorgan et al., 2011). The authors distinguished between two types of theories: (1) what they called "shallow" theories, which postulate either direct connectivity between modality-specific representations (i.e., distributed multimodal semantics) or a connectivity of those areas into a single mediating construct (a cross-modal semantic hub); and (2) "deep" theories, which postulate a hierarchy of mediating constructs (convergence zones), which progressively combine modality-specific knowledge into increasingly more cross-modal representations (higher order modality-specific areas, bi-modal areas, tri-modal areas, and so on). Note that this classification would include both the multimodal semantic models and the hub-and-spoke framework discussed above under the "shallow" classification. The assumption is that the two types of models make different predictions about the processing time required to integrate information coming from a single modality vs. the processing time required to ingrate information coming from multiple modalities. They used four verbal feature verification tasks and the results supported a deep model of semantics. One weakness of this study is the use of verbal stimuli. The unspoken assumption that these stimuli will in fact activate modality-specific representations (such as visual or functional), and only then propagate activation forward to convergence zones or any associative areas is not discussed and may in fact have distinct implications for the different theories. This combined with the lack of imagining data to complement the behavioral findings hinders the ability to make claims about the processing involved in the tasks: what modality-specific and association areas are involved, what are the temporal dynamics, and so on. Nonetheless, the investigation provides further support to the notion that the semantic system involves more than simply distributed modality-specific areas, and presents an interesting and useful approach to distinguishing between models of semantics.

Returning to the original objective of our study, the natural next step in establishing the structure of the semantic system is to inquire about the conceptual organization of cross-modal representations. We found that modality-specific pools of knowledge exhibit meaningful and unique structure. How does this structure compare to a cross-modal combined representation of this knowledge? And how do the various modalities contribute? A number of previous investigations (e.g., Andrews et al., 2009; Durda et al., 2009; Steyvers, 2010) seem to attribute a prominent role to verbal information – contrasting it with feature type information independent of its modality, as opposed to treating it as yet another modality of experience as we have done in our approach (see also, Plaut, 2002; Rogers et al., 2004, for similar approaches within a connectionist paradigm).

Last but not least, we will consider a few methodological issues and contributions from our current work. Feature listings have often been criticized in the past as an unreliable method to probe people's semantic knowledge (e.g., Murphy and Medin, 1985; Rogers et al., 2004) for at least three reasons: (1) participants know much *more* about each concept than what they list in any one study (therefore, the lists are incomplete); (2) the features that participants give is a potentially *random* sample of their full knowledge (therefore, the lists are inconsistent/variable); and (3) the knowledge is probed *verbally* for all types of features (therefore, the information provided is filtered by vocabulary demands which may impact some modalities of knowledge more than others because the attributes in that domain do not have verbal labels or are hard to express, e.g., elements of praxis or non-verbal auditory sounds).

We found that one way to improve the quality of the feature listings was to consider all features listed (independent of concept) and to re-score each concept against each feature. This guards against (quite common) cases where participants generate a certain feature (e.g., "has a long neck") for one concept (e.g., *swan*) but not for another (e.g., *peacock*), even though it applies to both. Also, features that describe identical or similar aspects of the concept (e.g., "has a box-like shape" and "looks like a square") should be grouped together to minimize redundancy and improve concept overlap. Although laborious (especially if undertaken for more than the 52 concepts considered in this study), combining these two techniques counteracts the incomplete and variable nature of feature listings. In this study, for example, it greatly improved the representational density of the Cree and McRae data set, which in turn resulted in an improved and more informative emergent structure.

The present investigation also provided an insight into the third concern listed above. We found that conceptual representations based on verbally reported features, taken to provide information about non-verbal modalities, are distinctly different from conceptual representations based on verbal experience (i.e., using the concept names in context). The established similarity and coherence between perceptual representations based on feature norms (re-scored) and perceptual representations based on participants' drawings further supported the notion that verbal feature listings provide a close approximation of modality-specific non-verbal knowledge.

Of course, we have only considered a handful of knowledge types. The fact is that experience in some modalities (olfactory, gustatory, tactile, etc.) may not be as easily verbalized as visual or motor experience. Some researchers have solved this problem by asking their participants to give a rating of how relevant each knowledge type is to a specific item, instead of listing features of various types (e.g., Gainotti et al., 2009; Hoffman and Lambon Ralph, 2012). Gainotti et al. (2009) presented college students with the pictures and names of 28 living things and 21 artifacts, and asked them to rate the familiarity of each item and to indicate (on a scale from 0 to 7) how relevant each source of modality-specific knowledge was in defining each item. The knowledge types included visual, auditory, tactile, olfactory, gustatory, motor/functional, and encyclopedic. The raw scores were transformed into percentages indicating the relative contribution of each modality to each concept. Their results indicated that olfactory and gustatory experience was significantly relevant only to the plant subcategory (fruits, vegetables, and flowers), whereas tactile experience was most relevant to the same categories as functional experience, namely artifacts such as tools, clothing, and furniture. Even though the clever methodology allowed the researchers to collect data for all modalities (even those where features or attributes may not be easy to report), the analysis suffers from the same assumption as the other studies discussed earlier – namely that there is categorical organization within each modality, which as we have established here is not the standard structure across modalities.

CONCLUSION

This study looked at three distinct methods of probing modality-specific knowledge (feature listings, drawings, and verbal co-occurrence statistics) to assess the conceptual structure in four modalities: perceptual, functional, encyclopedic, and verbal. Unlike previous studies, we did not assume that taxonomic categories exist in each knowledge type. Instead, we utilized a data-driven approach to reveal distinct and logical organization of concepts in each modality. Only the perceptual modality consistently exhibited significant categorical structure. Verbal representations had the most idiosyncratic organization, weakly related to the functional representations and very dissimilar from the perceptual and encyclopedic representations, which were somewhat similarly organized. Thus, the semantic system draws from these rich and multifaceted modality-specific pools of knowledge, each with a complex representational structure, to form coherent transmodal representations.

ACKNOWLEDGMENTS

This work was supported by an EPSRC (UK) institutional strategic award to the University of Manchester (EP/K005294/1) and an MRC (UK) programme grant to Matthew A. Lambon Ralph (MR/J004146/1).

REFERENCES

Acosta-Cabronero, J., Patterson, K., Fryer, T. D., Hodges, J. R., Pengas, G., Williams, G. B., et al. (2011). Atrophy, hypometabolism and white matter abnormalities in semantic dementia tell a coherent story. *Brain* 134, 2025–2035.

Andrews, M., Vigliocco, G., and Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychol. Rev.* 116, 463–498.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34, 149–159.

Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J., and Lambon Ralph, M. A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cereb. Cortex* 20, 2728–2738.

Binney, R. J., Parker, G. J. M., and Lambon Ralph, M. A. (2012). Convergent connectivity and graded specialization in the rostral human temporal lobe as revealed by diffusion-weighted imaging probabilistic tractography. *J. Cogn. Neurosci.* 24, 1998–2014.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., and Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia* 38, 1207–1215.

Chao, L. L., Haxby, J. V., and Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat. Neurosci.* 2, 913–919.

Cree, G. S., and McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *J. Exp. Psychol. Gen.* 132, 163–201.

Durda, K., Buchanan, L., and Caron, R. (2009). Grounding co-occurrence: identifying features in a lexical co-occurrence model of semantic memory. *Behav. Res. Methods* 41, 1210–1223.

Eggert, G. H. (1977). *Wernicke's Works on Aphasia: A Source-Book and Review*. The Hague: Mouton.

Gainotti, G., Ciaraffa, F., Silveri, M. C., and Marra, C. (2009). Mental representation of normal subjects about the sources of knowledge in different semantic categories and unique entities. *Neuropsychology* 23, 803–812.

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., and Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: analyses of the semantic attributes of living and nonliving concepts. *Cogn. Neuropsychol.* 18, 125–174.

Goldberg, R. F., Perfetti, C. A., and Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *J. Neurosci.* 26, 4917–4921.

González, J., Barros-Loscertales, A., Pulvermuller, F., Meseguer, V., Sanjuán, A., Belloch, V., et al. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage* 32, 906–912.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., and Smith, L. (2009). Categorical structure among shared features in networks of early-learned nouns. *Cognition* 112, 381–396.

Hodges, J. R., Patterson, K., Oxbury, S., and Funnell, E. (1992). Semantic dementia: progressive fluent aphasia with temporal lobe atrophy. *Brain* 115, 1783–1806.

Hoffman, P., and Lambon Ralph, M. A. (2012). Shapes, scents and sounds: quantifying the full multi-sensory basis of conceptual knowledge. *Neuropsychologia* 51, 14–25.

Hoffman, P., Lambon Ralph, M. A., and Rogers, T. T. (2012). Semantic diversity: a measure of semantic ambiguity based on variability in the contextual usage of words. *Behav. Res. Methods.* PMID:23239067. [Epub ahead of print].

Hoffman, P., Rogers, T. T., and Lambon Ralph, M. A. (2011). Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: insight using a novel method to quantify contextual variability in meaning. *J. Cogn. Neurosci.* 23, 2432–2446.

Kobayashi, M. (2006). Functional organization of the human gustatory cortex. *J. Oral Biosci.* 48, 244–260.

Lambon Ralph, M. A., and Patterson, K. (2008). Generalization and differentiation in semantic memory: insights from semantic dementia. *Ann. N. Y. Acad. Sci.* 1124, 61–76.

Lambon Ralph, M. A., Sage, K., Jones, R. W., and Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2717–2722.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., and Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* 38, 487–497.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., and Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* 270, 102–105.

Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature* 379, 649–652.

Mayberry, E. J., Sage, K., and Lambon Ralph, M. A. (2011). At the edge of semantic space: the breakdown of coherent concepts in semantic dementia is constrained by typicality and severity but not modality. *J. Cogn. Neurosci.* 23, 2240–2251.

McNorgan, C., Reid, J., and McRae, K. (2011). Integrating conceptual knowledge within and across representational modalities. *Cognition* 118, 211–233.

Moore, C. J., and Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain* 112, 943–962.

Murphy, G. L., and Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychol. Rev.* 92, 289–316.

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987.

Peelen, M. V., and Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *J. Neurosci.* 32, 15728–15736.

Perani, D., Schnur, T., Tettamanti, M., Gorno-Tempini, M., Cappa, S. F., and Fazio, F. (1999). Word and picture matching: a PET study of semantic category effects. *Neuropsychologia* 37, 293–306.

Plaut, D. C. (2002). Graded modality-specific specialization in semantics: a computational account of optic aphasia. *Cogn. Neuropsychol.* 19, 603–639.

Pobric, G., Jefferies, E., and Lambon Ralph, M. A. (2007). Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20137–20141.

Pobric, G., Jefferies, E., and Lambon Ralph, M. A. (2010). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Curr. Biol.* 20, 964–968.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and deterioration of semantic memory: a neurological and computational investigation. *Psychol. Rev.* 111, 205–235.

Rogers, T. T., and McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach.* Cambridge, MA: MIT Press.

Simmons, W. K., Martin, A., and Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cereb. Cortex* 15, 1602–1608.

Smith, E. E., and Medin, D. L. (1981). *Categories and Concepts.* Cambridge, MA: Harvard University Press.

Snowden, J. S., Goulding, P. J., and Neary, D. (1989). Semantic dementia: a form of circumscribed cerebral atrophy. *Behav. Neurol.* 2, 167–182.

Steyvers, M. (2010). Combining feature norms and text data with topic models. *Acta Psychol. (Amst.)* 133, 234–243.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: inferring "how" from "where." *Neuropsychologia* 41, 280–292.

Tranel, D., Damasio, H., and Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia* 35, 1319–1327.

Vandenberghe, R., Price, C., Wise, R., Josephs, O., and Frackowiak, R. S. J. (1996). Functional-anatomy of a common semantic system for words and pictures. *Nature* 383, 254–256.

Visser, M., Jefferies, E., Embleton, K. V., and Lambon Ralph, M. A. (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobe. *J. Cogn. Neurosci.* 24, 1766–1778.

Visser, M., and Lambon Ralph, M. A. (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *J. Cogn. Neurosci.* 23, 3121–3131.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Q. J. Exp. Psychol.* 27, 635–657.

Wittgenstein, L. (2001). *Philosophical Investigations: The German Text, with a Revised English Translation 50th Anniversary Commemorative Edition.* Oxford: Wiley-Blackwell.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 September 2012; paper pending published: 05 October 2012; accepted: 03 December 2012; published online: 31 December 2012.

Citation: Dilkina K and Lambon Ralph MA (2012) Conceptual structure within and between modalities. Front. Hum. Neurosci. ***6****:333. doi: 10.3389/fnhum.2012.00333*

APPENDIX

FULL LIST OF CONCEPTS AND MOST SIMILAR ITEMS IN EACH DATA SET

Procedure

In each data set, the cosine between each pair of representational vectors was computed. The items included in the following lists have a cosine with the target concept at least two standard deviations above the average cosine value for that target concept. NONE = no items exceeded the threshold. NA = this target concept was not present in the set.

Table A1 | Lists for the four sensory/perceptual sets and the verbal set.

Concept	Rogers et al. visual	Garrard et al. sensory	Cree and McRae perceptual	Cree and McRae edited perceptual	Hoffman et al. verbal
Airplane	Bus, lorry, train, helicopter	Helicopter, train, bus	Eagle, sled	Bus, helicopter, train, motorcycle	Helicopter, tiger
Alligator	Turtle, dog	Tiger, camel	Banana, rabbit, frog, turtle	Turtle	Turtle, elephant
Apple	Tomato, orange, banana, pineapple	Cherry, tomato, carrot, orange, banana, pineapple	Tomato, cherry, pineapple, orange	Tomato, cherry, pineapple, banana	Pineapple, carrot, banana
Axe	Screwdriver, paintbrush, hammer	Hammer, screwdriver	Hammer, screwdriver	Hammer, scissors, screwdriver	Hammer, sled
Banana	Pineapple, orange, apple, carrot	Tomato, apple, cherry, carrot	Tomato, apple, alligator, orange	Orange, apple, tomato, carrot	Pineapple, carrot, chicken
Barrel	Dustbin, envelope, banana, candle	Dustbin, axe	Dustbin, camel	Basket, dustbin	Rabbit
Basket	Toaster	Dustbin, hairbrush, stool	Paintbrush, hammer, hairbrush	Barrel, envelope, paintbrush, dustbin	Dustbin
Bike	Motorcycle, screwdriver	Motorcycle, sled, bus	Motorcycle, train, sled	Motorcycle, sled	Motorcycle, bus, lorry
Hairbrush	Paintbrush, screwdriver	Paintbrush, hammer, screwdriver	Paintbrush, sled, scissors, dustbin, screwdriver	Paintbrush, sled	Paintbrush, scissors
Bus	Lorry, train, airplane	Helicopter, airplane, train, motorcycle	Motorcycle, lorry, train	Train, airplane, helicopter, lorry, motorcycle	Bike, lorry
Camel	Cow, horse, dog	Cow, horse	Barrel, cow, horse	Horse, cow	Elephant, tiger
Candle	Pineapple, stool, carrot	Hairbrush	Envelope, paintbrush, pliers, sled	Pliers, envelope, paintbrush	Basket, stool
Carrot	Banana, pineapple, candle, cherry	Apple, cherry, tomato, orange, banana	Paintbrush, screwdriver, banana, alligator, orange, tiger	Orange, banana, cherry, apple, pineapple	Tomato, chicken, pineapple
Cat	Dog, squirrel, tiger	Mouse, squirrel, rabbit, dog	Horse, squirrel, dog, mouse, elephant	Dog, tiger	Dog, squirrel
Cherry	Pineapple, carrot, tomato	Tomato, apple, orange, carrot	Apple, tomato, dustbin	Apple, tomato, orange, pineapple	Apple, orange, toaster
Chicken	Peacock, owl, eagle, ostrich	Duck, eagle, owl	Eagle, duck, owl, swan	Peacock	Tomato, carrot
Cow	Camel, horse, dog	Horse, camel	Horse, elephant	Horse, camel	Elephant
Dog	Tiger, cat, cow	Horse, mouse, cat, squirrel	Cat, squirrel, horse, elephant	Cat	Cat, sled
Duck	Swan, chicken, peacock, ostrich	Chicken, penguin	Swan, penguin, eagle, chicken, owl	Swan	Turtle, swan, chicken, ostrich
Dustbin	Barrel, envelope, piano, candle	Barrel, basket	Hairbrush, stool, key, sled	Barrel, basket	Basket, cat, lorry
Eagle	Peacock, chicken, ostrich, owl	Owl, chicken, swan, ostrich	Chicken, duck, owl, swan, peacock	Owl	Owl
Elephant	Cow, dog, camel	Turtle	Horse, squirrel	Horse	Tiger, turtle, squirrel, alligator
Envelope	Barrel, dustbin	Bus, train	Candle, paintbrush, pliers, sled	Basket, dustbin	Dustbin, scissors, basket

(Continued)

Table A1 | Continued

Concept	Rogers et al. visual	Garrard et al. sensory	Cree and McRae perceptual	Cree and McRae edited perceptual	Hoffman et al. verbal
Frog	Rabbit, alligator, mouse	Alligator	Turtle, cat, cow	Turtle	Turtle, rabbit
Hammer	Screwdriver, key	Axe, hairbrush	Axe, basket, screwdriver	Axe, screwdriver, hairbrush	Screwdriver, pliers
Helicopter	Airplane, bus, lorry	Airplane, bus, train	Scissors, train, hammer	Airplane, bus, train	Airplane, motorcycle, tiger
Horse	Cow, camel	Mouse, dog	Elephant, squirrel, cat	Cow, camel	Bike
Key	Hammer, screwdriver	Pliers, hairbrush, hammer, screwdriver	Dustbin, stool	Dustbin, stool	Piano
Lorry	Bus, train	Motorcycle, airplane, train	Bus, motorcycle, train	Bus, train, airplane, motorcycle, helicopter	Motorcycle, bike, bus
Motorcycle	Bike	Bike, sled, lorry	Train, bus, bike, lorry	Bike, airplane, bus, train, lorry	Bike, lorry
Mouse	Rabbit	Rabbit, squirrel, cat, horse	Squirrel, cat, rabbit	Rabbit, squirrel	Paintbrush, rabbit, frog
Orange	Apple, banana, tomato	Cherry, tomato, apple, carrot	Tomato, apple, cherry	Banana, tomato, cherry, carrot, apple, pineapple	Pineapple, carrot
Ostrich	Swan, chicken, eagle, peacock	Eagle, peacock, swan	Eagle, swan, chicken	Peacock, eagle	Duck, swan, squirrel, alligator, turtle
Owl	Chicken, peacock, eagle, penguin	Eagle, chicken	Eagle, swan, chicken, duck	Eagle	Eagle
Paintbrush	Screwdriver, hairbrush, axe, hammer	Hairbrush, axe	Basket, hairbrush, screwdriver	Hairbrush, basket, sled	Hairbrush, mouse
Peacock	Chicken, eagle, owl, ostrich	Ostrich, chicken, duck, eagle, owl	Eagle, owl, swan	Ostrich, swan, duck	Swan, orange
Penguin	Owl, chicken	Duck, swan, chicken, eagle	Duck, swan	Duck, swan	Swan, cat
Piano	Dustbin, stool, sled	Barrel, hammer	Barrel, cow	Bus	Key, hammer
Pineapple	Tomato, banana, apple	Apple, cherry, tomato, carrot, banana, orange	Apple, orange	Apple, cherry, orange, tomato	Banana, carrot, tomato, chicken
Pliers	Scissors, orange, apple	Key, hammer, axe	Scissors	Hairbrush, sled	Screwdriver, hammer
Rabbit	Squirrel	Mouse, cat, squirrel	Mouse, alligator	Squirrel, mouse	Cat, squirrel, mouse
Scissors	Pliers, key, orange	Screwdriver, axe	Hairbrush, pliers, sled, helicopter, dustbin	Axe, hairbrush, pliers, sled	Hairbrush, pliers, screwdriver, hammer
Screwdriver	Paintbrush, hammer, axe	Axe, hairbrush, scissors	Hairbrush, hammer, paintbrush, axe	Hammer, paintbrush, axe	Pliers, hammer
Sled	Stool, piano, bus	Motorcycle, bike	Hairbrush, scissors	Hairbrush, paintbrush	Dog, cat
Squirrel	Rabbit, cat	Mouse, cat	Horse, mouse, elephant, cat	Rabbit, mouse, cat	Turtle, elephant, tiger
Stool	Sled, candle, piano	Basket, sled	Dustbin	Dustbin, hairbrush, sled	Candle
Swan	Ostrich, duck, chicken	Eagle, penguin, ostrich	Duck, penguin, eagle, owl, chicken	Duck, peacock	Duck, ostrich, pineapple
Tiger	Dog, cat	Mouse, squirrel, horse	Cat, eagle, piano, rabbit	Cat, dog	Elephant, turtle, squirrel
Toaster	Basket, envelope, bus	Lorry	Hairbrush, key, pliers, sled	Dustbin, helicopter, envelope	Cherry, apple, dustbin
Tomato	Apple, pineapple, orange	Cherry, apple, orange, banana	Apple, orange, cherry, banana	Apple, cherry, orange	Chicken, carrot, pineapple
Train	Bus, lorry, airplane	Airplane, helicopter, bus, lorry	Motorcycle	Bus, airplane, helicopter, motorcycle, lorry	Camel, sled
Turtle	Alligator	Elephant	Frog, stool, squirrel	Alligator	Elephant, squirrel, tiger, alligator

Table A2 | Lists for the three functional and the three encyclopedic sets.

Concept	Garrard et al. functional	Cree and McRae functional	Cree and McRae edited functional	Garrard et al. encyclopedic	Cree and McRae encyclopedic	Cree and McRae edited encyclopedic
Airplane	Helicopter, sled	Train, bus	Bus, helicopter, lorry	Helicopter, piano	Helicopter	Helicopter, bus, train
Alligator	Penguin, swan, turtle	NA	Frog, ostrich, owl, peacock, penguin, turtle	Tiger	Axe, frog, scissors	Turtle, frog, peacock, penguin
Apple	Cherry, tomato, banana	Cherry	Cherry, orange, pineapple, tomato	Orange, banana, carrot, pineapple, tomato	Cherry, banana, orange, pineapple, carrot	Cherry, carrot, banana, tomato
Axe	Scissors, pliers	Scissors	Scissors, hairbrush, paintbrush	Hammer, camel, ostrich	Alligator, motorcycle, tiger	Scissors, paintbrush
Banana	Carrot, cherry, apple, orange, tomato, pineapple	Chicken, duck, rabbit, tomato	Orange, pineapple, cherry	Pineapple, orange, apple, carrot, cherry	Cherry, apple, orange, pineapple	Pineapple, orange, apple
Barrel	Dustbin, basket	NONE	Dustbin, basket	NA	NA	Dustbin, hammer, pliers, scissors
Basket	Dustbin, barrel	NONE	Dustbin, barrel	NA	NONE	Barrel, dustbin, hammer, pliers, scissors
Bike	Sled, motorcycle	Camel, horse	Camel, motorcycle, sled, horse	Motorcycle, lorry, train	NONE	Motorcycle, lorry, paintbrush, sled
Hairbrush	Paintbrush, screwdriver, pliers	NA	Paintbrush, scissors	NONE	NA	Dustbin, paintbrush
Bus	Train, lorry, motorcycle	Airplane, train	Airplane, train, helicopter	Lorry, train	NONE	Lorry, motorcycle, paintbrush
Camel	Horse, cow	Bike, horse	Elephant, bike, stool, horse	Axe, ostrich, elephant	NONE	Ostrich, elephant
Candle	Toaster, basket, stool	NONE	Lorry, stool	NONE	NONE	Paintbrush, toaster
Carrot	Banana, pineapple, orange	Tomato	Tomato, orange, pineapple	Apple, banana, cherry, pineapple, orange, tomato	Apple, tomato	Apple, tomato
Cat	Squirrel	NA	Dog	Peacock, chicken	Cow, dog, chicken	Dog, mouse, horse
Cherry	Apple, tomato, banana, orange	Apple	Apple, banana, orange, pineapple	Orange, banana, carrot, pineapple, tomato	Apple, banana, orange, pineapple	Apple
Chicken	Owl, peacock, ostrich	Duck, rabbit, banana, tomato	Duck, cow, rabbit	Peacock, cow, cat	Cow, peacock	Cow, dog, horse
Cow	Mouse, rabbit	NONE	Rabbit, chicken, duck	Chicken, rabbit, horse	Chicken, cat, dog	Chicken, dog, horse
Dog	Horse	NONE	Cat, horse, tiger	Peacock, cat, chicken	Cat, cow, chicken	Cat, cow, chicken, horse
Duck	Swan, penguin, turtle, peacock	Chicken, elephant, banana, tomato	Chicken, rabbit	Turtle, rabbit	Swan, turtle, peacock	Swan, frog, turtle
Dustbin	Basket, barrel	NA	Basket, barrel	NA	NA	Hairbrush, barrel, hammer, pliers, scissors
Eagle	Owl, swan, peacock	NA	Ostrich, owl, peacock, penguin, swan, turtle	Owl, tiger	Turtle, duck, peacock, chicken, penguin, swan	Swan, peacock, turtle, duck
Elephant	Camel, horse, cow	Tiger, duck, rabbit	Camel, horse	Ostrich	Ostrich, tiger, peacock	Tiger, camel
Envelope	Paintbrush, screwdriver, basket, sled	NONE	Lorry, stool	NA	NONE	Paintbrush

(Continued)

Table A2 | Continued

Concept	Garrard et al. functional	Cree and McRae functional	Cree and McRae edited functional	Garrard et al. encyclopedic	Cree and McRae encyclopedic	Cree and McRae edited encyclopedic
Frog	Rabbit	NA	Alligator, ostrich, owl, peacock, penguin, turtle	Duck, penguin, turtle, alligator, swan	Turtle, alligator	Turtle, alligator, penguin
Hammer	Pliers, scissors, paintbrush, screwdriver, hairbrush	Pliers, screwdriver	Pliers, screwdriver, paintbrush, scissors	Axe, camel	Pliers, screwdriver	Screwdriver, barrel, dustbin, scissors
Helicopter	Airplane, eagle	Camel, airplane	Airplane, bus, train	Airplane, motorcycle	Airplane	Airplane
Horse	Dog, camel	Bike, camel	Camel, elephant, bike	Duck, peacock, cow	Cow, peacock, chicken, swan	Cow, chicken, cat, dog
Key	Basket, dustbin	NONE	Hairbrush, paintbrush	NONE	NONE	Barrel, dustbin, hammer, pliers, scissors
Lorry	Train, bus	NA	Airplane, train	Bus, train, motorcycle	NA	Bus, motorcycle
Motorcycle	Bus, bike, train, airplane	Camel, airplane, helicopter	Bike, camel, helicopter, sled, stool	Bike, lorry, train	Axe, scissors, sled	Lorry, bike, bus, axe
Mouse	Cow, tiger, squirrel, rabbit	NONE	Squirrel, alligator, rabbit	Squirrel, dog, tiger	Squirrel	Rabbit, cat
Orange	Banana, pineapple, carrot, cherry, apple, tomato	NONE	Banana, apple, tomato, carrot, cherry	Apple, banana, cherry, pineapple	Cherry, apple, banana, pineapple	Banana, pineapple
Ostrich	Peacock	NA	Eagle, alligator, frog, swan	Camel, elephant	Peacock, elephant	Camel, turtle, peacock
Owl	Eagle, swan, chicken, peacock	NA	Eagle, alligator, frog, swan	Eagle, tiger	Squirrel	Squirrel, eagle, rabbit
Paintbrush	Screwdriver, hairbrush, pliers	NONE	Scissors, hairbrush, pliers	NONE	NA	Axe, hairbrush, piano
Peacock	Duck, swan	NA	Eagle, alligator, frog, swan	Cat, chicken, dog, horse, duck	Ostrich, chicken, penguin, swan	Turtle
Penguin	Swan, alligator, duck	NA	Eagle, alligator, frog, swan	Frog, duck, turtle, alligator, swan	Peacock, chicken, ostrich	Turtle
Piano	Stool, dustbin	NONE	Bike, lorry, motorcycle, sled, stool	Airplane	NONE	Paintbrush
Pineapple	Carrot, orange, banana, cherry	NONE	Banana, apple, tomato, carrot, cherry	Banana, orange, tomato, apple, carrot, cherry	Cherry, apple, banana, orange	Banana, orange
Pliers	Screwdriver, scissors, paintbrush, hammer	Screwdriver, hammer, horse	Hammer, screwdriver, paintbrush, scissors	NONE	Hammer, screwdriver	Screwdriver, barrel, dustbin, scissors
Rabbit	Cow, frog, tiger, squirrel	Chicken, elephant, banana, tomato	Cow, duck, alligator, chicken	Duck, cow	NONE	Squirrel, mouse, peacock
Scissors	Pliers, paintbrush, screwdriver, axe, hammer, hairbrush	Axe	Paintbrush, hairbrush, pliers	NONE	Alligator, motorcycle, tiger	Axe
Screwdriver	Paintbrush, pliers, hairbrush	Pliers, hammer	Pliers, hammer, paintbrush, scissors	NONE	Hammer, pliers	Hammer, pliers
Sled	Bike, airplane, camel	NONE	Bike, motorcycle, stool	NONE	Motorcycle	Paintbrush, bike

(Continued)

Table A2 | Continued

Concept	Garrard et al. functional	Cree and McRae functional	Cree and McRae edited functional	Garrard et al. encyclopedic	Cree and McRae encyclopedic	Cree and McRae edited encyclopedic
Squirrel	Cat, mouse, rabbit	NA	Mouse, alligator, rabbit	Turtle, rabbit	Owl, mouse	Rabbit, owl
Stool	Piano, toaster, basket, candle, dustbin	NONE	Camel	NA	None	Paintbrush, toaster
Swan	Penguin, duck, alligator	NA	Eagle, ostrich, owl, peacock, penguin, turtle	Alligator	Duck, peacock	Duck, eagle, peacock
Tiger	Mouse, rabbit, cat, squirrel	Elephant	Eagle, alligator	Alligator, eagle, owl	Elephant, axe, scissors, ostrich	Elephant
Toaster	Candle, basket, stool, dustbin	NONE	Lorry, stool	NONE	NONE	Paintbrush, stool
Tomato	Cherry, apple, banana	Chicken, duck, rabbit, banana	Carrot, orange, pineapple, apple	Pineapple, apple, banana, carrot, cherry	Carrot	Carrot, apple, paintbrush
Train	Bus, lorry, motorcycle	Airplane, bus	Bus, helicopter, lorry	Lorry, bus, motorcycle	NONE	Bus, paintbrush
Turtle	Duck, alligator, penguin	NA	Eagle, alligator, frog, swan	Duck, squirrel, ostrich	Frog	Peacock, alligator, penguin, ostrich

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 18 April 2012
doi: 10.3389/fnhum.2012.00085

Apples are not the only fruit: the effects of concept typicality on semantic representation in the anterior temporal lobe

Anna M. Woollams*

Neuroscience and Aphasia Research Unit, School of Psychological Sciences, University of Manchester, Manchester, UK

Edited by:
Penny M. Pexman, University of Calgary, Canada

Reviewed by:
Chris McNorgan, Northwestern University, USA
Sebastian Crutch, University College London, UK

****Correspondence:***
Anna M. Woollams, Neuroscience and Aphasia Research Unit, School of Psychological Sciences, Zochonis Building, University of Manchester, Brunswick Street, Manchester M13 9PL, UK.
e-mail: anna.woollams@manchester.ac.uk

Intuitively, an apple seems a fairly good example of a fruit, whereas an avocado seems less so. The extent to which an exemplar is representative of its category, referred to here as concept typicality, has long been thought to be a key dimension determining semantic representation. Concept typicality is, however, correlated with a number of other variables, in particular age of acquisition (AoA) and name frequency. Consideration of picture naming accuracy from a large case-series of semantic dementia (SD) patients demonstrated strong effects of concept typicality that were maximal in the moderately impaired patients, over and above the impact of AoA and name frequency. Induction of a temporary virtual lesion to the left anterior temporal lobe, the region most commonly affected in SD, via repetitive Transcranial Magnetic Stimulation produced an enhanced effect of concept typicality in the picture naming of normal participants, but did not affect the magnitude of the AoA or name frequency effects. These results indicate that concept typicality exerts its influence on semantic representations themselves, as opposed to the strength of connections outside the semantic system. To date, there has been little direct exploration of the dimension of concept typicality within connectionist models of intact and impaired conceptual representation, and these findings provide a target for future computational simulation.

Keywords: concept typicality, picture naming, semantic dementia, repetitive transcranial magnetic stimulation, age of acquisition, frequency

The purpose of semantic knowledge is to allow us to recognize different instances of a particular item (e.g., that an apple and an orange are both fruits) and generalize to novel instances of that item irrespective of perceptual variation (e.g., that an avocado is also a fruit). Semantic representations are formed by extracting coherent covariation among conceptual features (i.e., seeds), allowing abstraction from the perceptual features (e.g., sweet, juicy). Of course, the perceptual features of a concept will be activated along with its semantic representation, but these are stored in separate regions of the brain. This notion of a combination of modality independent and modality-specific features that participate in semantic processing forms the basis of the hub and spoke model of semantic memory reproduced in **Figure 1** (Patterson et al., 2007). Within this model, semantic representations are stored in the anterior temporal lobe (ATL) that provides the amodal hub to link the modality-specific spokes of perceptual knowledge.

This model has gained considerable support from the study of patients with semantic dementia (SD), a progressive degenerative disease involving atrophy and hypometabolism of the ATLs, resulting in a selective and worsening deficit of semantic memory (Gorno-Tempini et al., 2011). These patients show deficits that cut across all modalities of testing, affecting their ability to name pictures; understand spoken words; recognize pictures, written words, environmental sounds and characteristic smells; and draw and use objects (Lambon Ralph et al., 2008). The pan-modal nature of the semantic deficits seen in SD argues strongly for the existence of amodal semantic representations housed in the ATLs, although this view is not universally accepted (Martin, 2007). It is certainly true that atrophy does spread beyond anterior temporal regions as SD progresses (Rohrer et al., 2009; Gordon et al., 2010), however evidence of the amodal nature of semantic representations in the ATLs has also been found using both distortion corrected functional imaging (Visser et al., 2010) and virtual lesion methodology using repetitive Transcranial Magnetic Stimulation (rTMS) (Pobric et al., 2010a).

To return to the apple and the avocado, it is clear that an apple is a good example of a fruit because it has features commonly found in other fruits (e.g., seeds, skin, sweet, juicy), whereas an avocado has fewer of these (e.g., seeds, skin) and also involves some more idiosyncratic features not found in many other fruits (e.g., savory, oily). The difference between the apple and the avocado can be seen as one of the typicality of their features within the fruit category. Since the proposal of the very first cognitive models of semantic memory, it has been recognized that the typicality of an item is a key dimension of its semantic representation. Smith et al. (1974) reported evidence of a concept typicality effect such that people are faster to verify that "An apple is a fruit" than "An avocado is a fruit." Indeed, concept typicality ratings confirm that an apple is judged to be a more typical fruit than an avocado

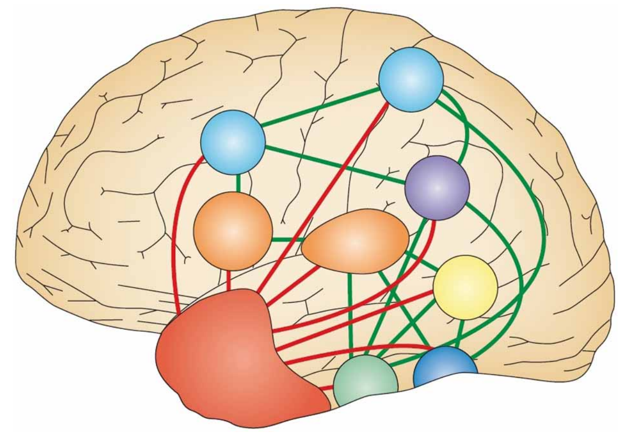


FIGURE 1 | The hub and spoke model of semantic processing. The various different modality-specific surface representations correspond to motion (yellow), colors (dark blue), shape (green), names (orange), actions (light blue) and task (purple), and are directly connected via green lines. These all connect (shown as red lines) to a shared, amodal "hub" (shown as a red area) in the anterior temporal lobes. At the hub stage, therefore, associations between different pairs of attributes (such as shape and name, shape and action, or shape and color) are all processed by a common set of neurons and synapses, regardless of the task. Adapted from Figure 1B of Patterson et al. (2007, p.977).

(Morrow and Duffy, 2005), given the latter's savory flavor and oily texture. Such findings motivated Collins and Loftus (1975) to incorporate spreading activation into their semantic network and indeed form the basis of the prototype theory of semantic representation proposed by Rosch and Mervis (1975).

These initial cognitive models of semantic memory are not dissimilar to more recent connectionist conceptualizations of semantic memory for concrete concepts, which incorporate the same characteristics of featural representations, spreading activation, and semantic distance (e.g., Plaut, 1996; McRae et al., 1997, 1999; Rogers et al., 2004; Dilkina et al., 2008, 2010). McRae et al. presented models in which the written word-form activated semantic representations that were made up of transparent features listed by human participants, and the connections between them were indexed to the extent to which a given feature pair co-occurred across the description of all items (e.g., seeds and sweet co-occur together in more concepts than seeds and savory). In this framework, McRae et al. were concerned with the typicality of features, rather than concepts: seeds is a typical feature of apple because apple contains other features that often co-occur with seeds (e.g., sweet and juicy), whereas it is an atypical feature of an avocado because avocado contains other features that do not often co-occur with seeds (e.g., savory and oily). Nevertheless, these models illustrate an important aspect of typicality, which is that it is determined not by the number or frequency of features, but rather by the intercorrelation between those features. In the case of concept typicality, an apple's features are highly intercorrelated with those of many other fruits, hence it represents a typical exemplar of the category, whereas an avocado's features are less intercorrelated with others in the category, hence it represents a more atypical exemplar.

The weakly intercorrelated features that characterize the representation of atypical items would be expected to be more vulnerable to damage than the strongly intercorrelated features of typical items. Hence there is a clear prediction that amongst those who suffer from a deficit in semantic memory arising from damage to the ATLs, atypical concepts like avocado will suffer more than typical ones like apple. Woollams et al. (2008) provided the first assessment of this prediction in a large case-series study of picture naming in SD, and found an overwhelmingly strong effect of typicality upon picture naming accuracy. Moreover, the patients' errors of commission were increasingly more typical than the target (e.g., an avocado might initially be called a mango and then later a pear). This influence of typicality upon performance in SD has since been confirmed in a receptive matching to sample task (Lambon Ralph et al., 2010; Mayberry et al., 2011).

There are, however, a number of other factors that may affect the susceptibility of items to the progressive semantic damage that defines SD (Lambon Ralph et al., 1998; Woollams et al., 2008). One is Age of Acquisition (AoA), which has been shown to correlate with typicality—more typical exemplars like apple tend to be learnt earlier than less typical ones like avocado (Holmes and Ellis, 2006). Another is name frequency, as more typical items like apple have names that are used more often than less typical ones like avocado. The goal of this paper is to explore the extent to which concept typicality uniquely affects the nature of the meaning we have in mind. This will be achieved in two ways: firstly, by considering its impact upon picture naming accuracy across the range of severity in SD, and secondly, by exploring how it responds to a temporary virtual lesion to the left ATL (lATL) induced by offline rTMS.

THE IMPACT OF ATL ATROPHY ON PICTURE NAMING

METHODS

Participants

The present data set was derived from all patients listed in the Cambridge MemBrain patient database who had a clinical diagnosis of SD and who had completed picture naming on at least one occasion. This yielded a total pool of 225 observations from 78 patients, collected between 1991 and 2006. Demographic and background neuropsychological data for five severity groups containing 45 observations each can be found in **Table 1**. These five severity groups were created on the basis of overall picture naming accuracy and are provided for expository purposes to illustrate the key features of this patient group over the course of the disease. These data demonstrated selective and progressive semantic deficit that characterizes SD: MMSE (Folstein et al., 1975) scores were universally low; nonverbal intelligence/problem solving remained high as indicated by stable performance on Raven's Coloured Progressive Matrices (Raven, 1962); visuospatial processing was preserved, as seen by normal scores on the Rey Immediate Copy Test (Lezak, 1976); normal performance on the Delayed Recall version indicated preserved episodic memory; and normal forward and backward digit span (Wechsler, 1987) scores showed preserved short term memory function. These contrast with the impaired and worsening performance apparent on both verbal and nonverbal measures of comprehension [the Cambridge Spoken Word Picture Matching Test (Bozeat et al.,

Table 1 | Demographic information and neuropsychological test scores associated with each of the 225 observations of picture naming data from semantic dementia patients included in the present study, grouped according to level of severity (Reproduced from Table 1 of Woollams et al. (2008, p.2505).

Group*		Age	Educ'n	MMSE (%)	Ravens (%)	Rey copy (%)	Rey delayed recall (%)	Digit span F	Digit span B	PPT pictures (%)	S-WPM (%)	Picture naming (%)
Mild	Mean	63	13	**88**	86	93	47	7	5	**89**	**96**	**85**
	SD	7	4	13	18	10	21	1	2	9	6	8
	N	45	45	39	16	42	33	39	39	36	44	45
Mild-moderate	Mean	61	12	**81**	77	89	36	6	4	**82**	**86**	**56**
	SD	7	2	15	14	16	21	1	1	11	10	12
	N	45	45	43	12	43	33	39	38	36	45	45
Moderate	Mean	63	11	**70**	82	89	33	6	4	**69**	**70**	**27**
	SD	8	3	17	19	16	18	2	1	15	15	5
	N	45	45	43	14	43	27	40	38	33	43	45
Moderate-severe	Mean	63	12	**62**	81	85	26	6	4	**68**	**52**	**13**
	SD	7	3	19	15	21	17	1	1	13	19	3
	N	45	42	35	17	42	32	39	38	34	43	45
Severe	Mean	64	12	**52**	69	85	24	6	4	**68**	**39**	**4**
	SD	6	3	19	20	18	13	1	1	13	27	3
	N	45	44	34	15	40	21	40	36	29	41	45

**Severity was determined on the basis of picture naming scores, divided into five groups with 45 observations per group.*

Figures in bold indicate performance more than two standard deviations below the control mean.

MMSE, mini-mental state examination; S-WPM, spoken word-picture matching; PPT, pyramids and palm trees test; F, forward; B, backward.

2000) and the picture version of the Pyramids and Palm Trees Test (Howard and Patterson, 1992)] and most prominently, the striking anomia apparent on the Cambridge Picture Naming Test.

Stimuli

To assess the impact of concept typicality upon naming performance in SD, the objects in the 48 or 64 item Cambridge Picture Naming Tests were subdivided into equal-N High and Low Typicality sets, based upon scores for these items in the Morrow and Duffy (2005) ontological concept typicality norms for a group of older adults on a scale from one (low typicality) to seven (high typicality). Typicality scores were not available for all items: for the 48-item version, there were 40 scores available, and for the 64-itemversion, there were 50 scores available. When combining across tests, 20 items appeared in both, and hence there was a total of 70 unique items, yielding 35 lower typicality items (mean = 5.46, SD = 0.60) and 35 higher typicality items (mean = 6.55, SD = 0.27). A full listing of these items is provided in the **Table A1**.

RESULTS

In order to assess the significance of this typicality effect, the 225 observations for the higher and lower typicality items were analyzed according to their severity group (1–15) in a 2 (typicality) by 15 (severity) ANOVA with typicality as a within-participants factor and severity as a between participants factor. Severity groups based upon overall naming accuracy were created rather than using severity as a continuous predictor as the use of groups allowed a parallel analysis of the impact of typicality and severity across items. Fifteen groups of fifteen observations were used as this balanced the need to keep group sizes reasonable whilst at the same time keeping group variation low. The by-participants analysis revealed strong main effects of typicality ($F_{(1, 210)} = 1079.19$, $p < 0.0001$), severity ($F_{(14, 210)} = 1218.08$, $p < 0.0001$) and their interaction ($F_{(14, 210)} = 1181.76$, $p < 0.0001$). A parallel analysis was conducted where the accuracy for each group of 15 participants was averaged across the 40 higher and 40 lower typicality items, allowing a 2 (typicality) by 15 (severity) ANOVA with typicality as a between items factor and severity as a within items factor. This confirmed the significant main effects of typicality ($F_{(1, 68)} = 11.43$, $p = 0.001$), severity ($F_{(5.40, 367.08)} = 272.99$, $p < 0.0001$)[1], and their interaction ($F_{(5.40, 367.08)} = 4.35$, $p = 0.001$). As can be seen in **Figure 2**, the concept typicality effect is most pronounced for the moderately impaired patients. Nonetheless, the typicality effect is significant at every level of severity.

Yet in this set of 70 unique items, the correlation between concept typicality and AoA (taken from the Morrow and Duffy (2005) ratings from older adults) was −0.479, $p < 0.001$ and between concept typicality and CELEX (Baayen et al., 1993) spoken frequency [from the N-watch database (Davis, 2005)] was 0.239, $p = 0.047$. As indicated by the correlation of typicality with AoA and frequency, these items were not selected to provide a controlled assessment of concept typicality effects. One way to determine the unique contribution of concept typicality to naming performance in SD is to control for AoA and frequency in the items analysis that considered severity. Item values for each variable were, therefore, entered as a linear predictor of the item accuracy for each of the 15 severity groups. The results

[1]Greenhouse-Geisser corrected values are provided to correct for non-sphericity.

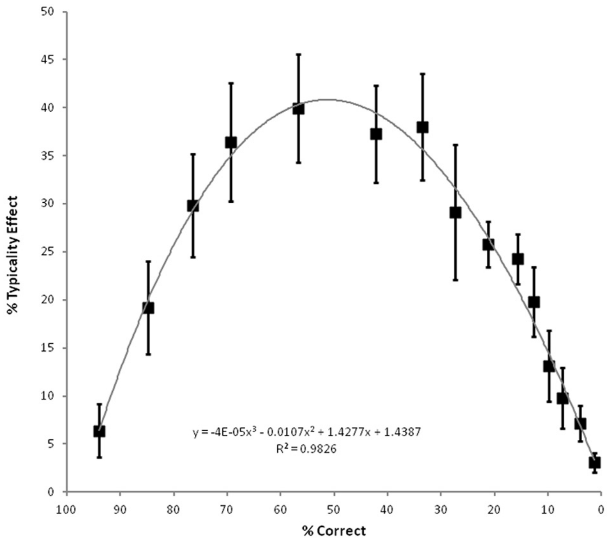


FIGURE 2 | Average magnitude of the concept typicality effect in picture naming accuracy (high typicality minus low typicality) for the 15 severity groups of 15 observations each, according to average overall level of accuracy, with cubic regression line and associated equation and fit. Error bars represent ±90% confidence intervals, hence those points in which they do not encompass the origin correspond to significant effects at an alpha level of 0.05 using a paired one-tailed *t*-test. Adapted from Figure 2 of Woollams et al., (2008, p.2507).

revealed significant main effects of frequency ($F_{(1, 66)} = 5.08$, $p = 0.028$) and AoA ($F_{(1, 66)} = 10.20$, $p = 0.002$) but not concept typicality ($F_{(1, 66)} = 2.39$, $p = 0.127$). Critically, however, the significant interaction between concept typicality and severity ($F_{(5.63, 371.82)} = 3.11$, $p = 0.007$) remained. An interaction between AoA and severity ($F_{(5.63, 371.82)} = 2.51$, $p = 0.023$) was also apparent, but the interaction between severity and frequency did not reach significance ($F_{(5.63, 371.82)} = 1.69$, $p = 0.151$). The results of this analysis show that concept typicality exerts an appreciable effect upon SD naming performance over and above AoA and frequency, particularly in patients with a moderately severe semantic deficit.

DISCUSSION

The present results revealed strong effects of concept typicality upon picture naming performance in SD, with the lower typicality items being most vulnerable to semantic damage, as expected. The largest impact of concept typicality was observed in the moderately severe SD patients, and this effect survived controlling for the correlated dimensions of AoA and frequency. The nonlinear relationship of severity to the size of the concept typicality effect indicates that initially it is the lower typicality items that decline most rapidly, and as the disease progresses, the decline in performance for higher typicality items accelerates until both are severely impaired. This result is consistent with the notion that concept typicality affects the nature of central amodal semantic representations housed in the ATLs that are damaged by progressive atrophy in SD, because the representations of lower typicality items are more vulnerable to semantic damage by virtue of their more idiosyncratic features. This interpretation produces the clear prediction that it should be possible to enhance the concept typicality effect seen in picture naming in normal participants via application of a temporary virtual lesion using rTMS to the lATL, the region most reliably affected in SD (Rohrer et al., 2008) and associated with the strongest levels of anomia (Lambon Ralph et al., 2001).

THE IMPACT OF lATL rTMS ON PICTURE NAMING

METHODS

Participants

Sixteen individuals participated in this picture naming experiment. All were students from the University of Manchester and participated in the study in exchange for £20. All spoke English as a first language and had normal or corrected-to-normal vision; three participants were male. The laterality quotient yielded by the Edinburgh Handedness Inventory (Oldfield, 1971) was 44.17 points on average (SD = 69.44). The mean age of the participants was 20.5 years old (SD = 2.84). None of the participants were taking medication and all were free from any history of neurological disease or mental illness.

Stimuli

The study used a repeated measures design in which participants named a set of 96 pictures before the application of rTMS and then another set of 96 pictures after the application of TMS, with the order of sets counterbalanced according to the order of enlistment. Ninety-three of these items had concept typicality ratings in the Morrow and Duffy (2005) norms for a group of younger adults. Of these, the 40 items with the highest and the 40 items with the lowest typicality ratings were selected for consideration. The mean ratings for these items on a variety of dimensions are provided in **Table 2**. Between items *t*-tests confirmed that the high and low typicality items differed significantly in their typicality ratings and also in their rated AoA and spoken frequency ($ts_{(1, 78)} > 2.43$, $ps < 0.017$). High and low typicality items did not differ significantly in terms of their visual complexity, name agreement or number of phonemes ($ts_{(1, 78)} < 1.19$, $ps > 0.237$). A full listing of these items is provided in the **Table A1**.

Table 2 | Means and standard deviations on a range of stimulus properties for the 40 low and 40 high typicality pictures used in the lATL rTMS study.

	Low typicality		High typicality	
	Mean	**SD**	**Mean**	**SD**
Typicality rating[1]	4.323975	0.90932	6.63975	0.248268
Age of acquisition[1]	3.044775	0.656757	2.327925	0.59091
Frequency per million[2]	4.962	9.432137	22.42325	44.39791
Visual complexity[3]	2.949	0.709672	2.7585	0.720682
Name agreement[3]	0.98375	0.02993	0.98425	0.024167
Number of phonemes[2]	4.625	1.496791	4.25	1.69085

[1] *Taken from Morrow and Duffy (2004) younger norms.*
[2] *Taken from information provided in the CELEX database (Baayen et al., 1993).*
[3] *Taken from Morrison et al. (1997) norms.*

Procedure

The DMDX experimental software package (Forster and Forster, 2003) was used to record RTs, vocal responses, and to display instructions and stimuli. Responses were collected by a voice key plus headset connected to an IBM compatible Pentium III computer with a 60 Hz refresh rate at 1280 by 1024 pixel screen resolution. Vocal responses were recorded from the beginning of the trial for a period of 1000 ms after the voice key triggered. Order of trial presentation was randomized anew for each participant within each block, and stimuli were presented in white on a black background. Mispronunciations and measurement errors were recorded by hand. Participants were instructed to name the centrally presented pictures as rapidly and accurately as possible. Trials began with a 500 ms fixation cross followed by the picture, which disappeared from the screen upon response or after 2000 ms. Each block of 96 pictures took around 5 min to complete.

The study used the virtual lesion method in which there was (a) a naming task (baseline), then (b) rTMS stimulation, and immediately after (c) an analogous naming task (probe). This meant that rTMS was delivered without a concurrent task and that the probe task was performed during the rTMS refractory period, which has been estimated as lasting for approximately 20 min (see Pobric et al., 2007). Focal magnetic stimulation was delivered using a 70 mm figure-of-eight coil attached to a MagStim Rapid2 stimulator (Magstim). Before experimental stimulation, motor threshold (MT) was determined for every participant as a visible twitch in the relaxed contralateral abductor pollicis brevis muscle in three out of six trials. Stimulation intensity for the experiment was then set at 120% of MT for each participant which resulted in an average of 62.42% (SD = 2.27) of the stimulator maximum output, and consisted of 10 min of 1 Hz stimulation. A structural T1-weighted MRI scan was acquired for each participant to guide TMS stimulation. As per Pobric et al. (2007), the ATL site was defined as the region 10 mm posterior from the tip of the left temporal lobe along the middle temporal gyrus, corresponding to the MNI co-ordinates of −53, 4, −32 (see **Figure 3**). For stimulation, this site was determined by co-registering cortical surface with 11 anatomical landmarks (inion, tip of the nose, left/right ear canals and left/right ear projections) some of which were marked prior to the scan with oil capsules (vertex, nasion, left/right ear tragus, and beneath lip in chin indentation). Co-registration was made using Ascension miniBIRD magnetic tracking system and MRIreg software (www.sph.sc.edu/comd/rorden/mrireg.html).

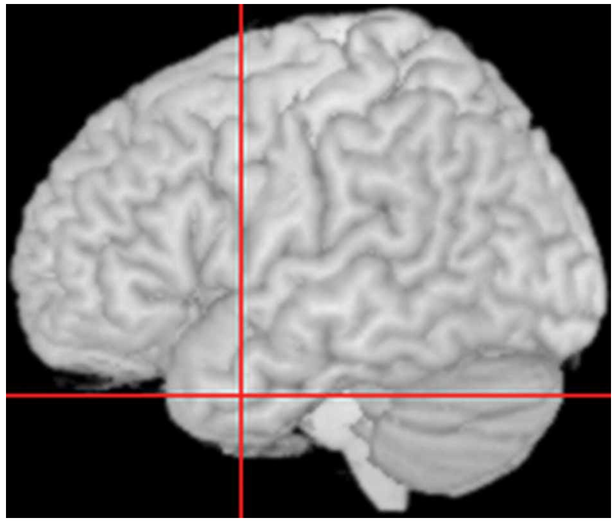


FIGURE 3 | Location of the left Anterior Temporal Lobe site to which repetitive Transcranial Magnetic Stimulation was applied. Crosshairs represent site of stimulation, corresponding to the co-ordinates −53, 4, −32 in MNI space.

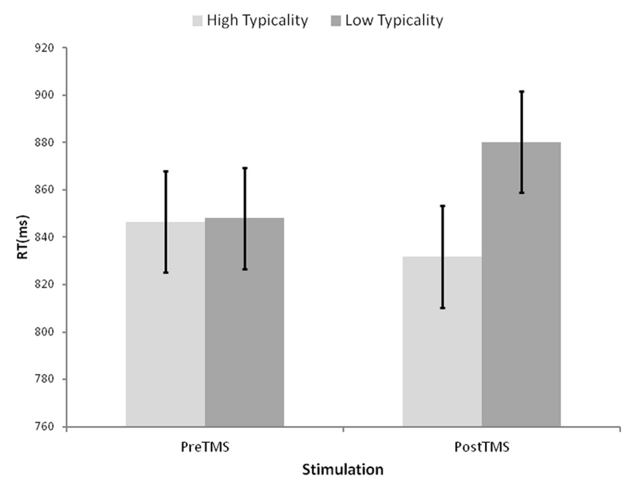


FIGURE 4 | Average reaction times for naming of high and low typicality pictures before and after the application of repetitive Transcranial Magnetic Stimulation to the left Anterior Temporal Lobe. Error bars represent ± within-participants 90% confidence intervals computed according to Loftus and Masson (1994, Equation 4, p. 485).

RESULTS

The correct reaction times and error rates were analyzed by two (typicality) by two (rTMS) ANOVAs, with typicality considered a within participants and between items variable, and rTMS considered as a within participants and within items variable. Analysis of the RT data, displayed in **Figure 4**, revealed a non-significant main effect of rTMS ($F_{1(1,\ 15)} = 0.51$, $p = 0.486$; $F_{2(1,\ 78)} = 0.81$, $p = 0.371$) and a main effect of typicality reliable by participants but not by items ($F_{1(1,\ 15)} = 5.48$, $p = 0.033$; $F_{2(1,\ 78)} = 0.78$, $p = 0.389$), which were qualified by an reliable interaction between the two, albeit it of marginal significance by participants ($F_{1(1,\ 15)} = 3.67$, $p = 0.075$; $F_{2(1,\ 78)} = 9.99$, $p = 0.002$). Follow-up one-tailed t-tests revealed that while the typicality effect was not significant prior to rTMS ($t_{1(15)} = 0.11$, $p = 0.480$; $t_{2(78)} = .34$, $p = 0.368$), it was after rTMS ($t_{1(15)} = 2.74$, $p = 0.008$; $t_{2(78)} = 1.90$, $p = 0.031$). Similarly, the effect of rTMS for low typicality items was significant ($t_{1(15)} = -1.84$, $p = 0.043$; $t_{2(39)} = -2.64$, $p = 0.006$) whereas that for high typicality items was not reliable ($t_{1(15)} = 0.85$, $p = 0.204$; $t_{2(39)} = 1.77$, $p = 0.042$). The parallel analyses of error rates revealed no reliable effects by participants ($F_{(1,\ 15)} < 2.50$, $p > 0.135$), and only a significant main effect of rTMS by items ($F_{(1, 78)} = 4.01$, $p = 0.049$, all other effects $F_{(1, 78)} < 1.44$, $p > 0.233$).

Across this set of 80 items, the correlation between concept typicality and AoA [taken from the Morrow and Duffy (2005) ratings from younger adults] was -0.498, $p < 0.0005$ and between concept typicality and CELEX (Baayen et al., 1993) spoken frequency [from the N-watch database (Davis, 2005)] was 0.261, $p = 0.019$. To determine the key stimulus dimension that was interacting with rTMS, item values for each variable were, therefore, entered as a linear predictor of the item RT and error rate pre and post rTMS. The results revealed a marginally significant main effect of typicality ($F_{(1, 76)} = 3.44$, $p = 0.068$), a significant main effect of AoA ($F_{(1, 76)} = 24.43$ $p < 0.0001$), but no reliable effect of frequency ($F_{(1, 76)} = 0.74$, $p = 0.392$). There was, however, a significant interaction between typicality and rTMS ($F_{(1, 76)} = 5.79$, $p = 0.019$), but not between AoA and rTMS ($F_{(1, 76)} = 0.21$, $p = 0.647$) or frequency and rTMS ($F_{(1, 76)} = 0.58$, $p = 0.448$). The main result of this analysis is that it is low typicality concepts that are particularly vulnerable to disruption due to lATL rTMS, again consistent with the notion that concept typlicality affects semantic representations housed in this area.

DISCUSSION

This study is the first to demonstrate a selective effect of lATL rTMS on the naming of low relative to high typicality concepts. In a previous study considering the impact of lATL stimulation on picture naming, Pobric et al. (2007) found a selective effect of lATL rTMS on naming at the specific level, in line with the greater deficits seen for this level in both SD patients (Adlam et al., 2006) and recent connectionist models of semantic representation (Rogers et al., 2004). Interestingly, Pobric et al. (2007) did not obtain an effect of lATL rTMS on basic level naming, in contrast to a later study (Pobric et al., 2010b). The stronger effect of lATL rTMS on low than high typicality concepts observed here suggests that the concept typicality of the stimuli to be named is a critical property to consider, and previous inconsistencies in the effects of lATL rTMS on basic level naming may have resulted from variation on this dimension.

The observed interaction between concept typicality and rTMS not only survived statistical control for the correlated variables of AoA and name frequency, but strikingly, neither of these variables interacted with rTMS. Although AoA has been shown to be a strong determinant of picture naming accuracy in SD (Lambon Ralph et al., 1998; Woollams et al., 2008), and frequency influences performance in SD across a variety of expressive and receptive tasks (Patterson et al., 2006; Caine et al., 2009; Jefferies et al., 2009), the impact of ATL rTMS on these effects has yet to be directly investigated. The present results demonstrate that future investigations of the impact of AoA and frequency on semantic representations in the ATL will need to take into account the correlated dimension of concept typicality.

GENERAL DISCUSSION

The present findings have shown a unique effect of concept typicality in picture naming when the function of the ATL is compromised. For picture naming in SD, the lower typicality concepts were those most vulnerable to damage across the full range of severity. Performance on lower typicality concepts also declined more rapidly, producing the largest typicality effect in picture naming accuracy for the moderately impaired patients. For normal picture naming, the lower typicality concepts were those most susceptible to disruption via offline rTMS of the lATL, the region that is most commonly affected in SD (Rohrer et al., 2008) and is associated with higher levels of anomia (Lambon Ralph et al., 2001). rTMS produced a significant effect of concept typicality in RTs that was not apparent in the baseline picture naming performance of normal participants for the same items.

The concept typicality effects observed here under conditions of semantic disruption are in accordance with the predictions of the early prototype theory of meaning representation proposed by Rosch and Mervis (1975): an apple is a closer to the prototype fruit than an avocado, hence it enjoys a processing advantage. Recent connectionist models of meaning representation implement many aspects of these early accounts, such as featural representations, spreading activation and semantic distance [see Dilkina et al. (2010) and Plaut (1996)]. Particularly relevant to the present result is a simulation reported by Rogers and McClelland (2004, p.203) within a connectionist model of concrete concepts that is able to generate names for items in response to activation of their visual features via a distributed semantic system. As these semantic representations are learnt in the course of mapping inputs to outputs, they bear an opaque relationship to those features found in empirically derived feature norms. Nevertheless, typicality can be quantified for each item in the model by means of computing the similarity of its representation with that of an averaged representation for the category (the group centroid). A strong relationship was found between the typicality of an item and the number of epochs taken to generate its basic level name. To date, however, the impact of damage to the semantic level upon the magnitude of this typicality effect in naming remains unexplored.

As mentioned earlier, McRae et al. (1997, 1999) used empirically derived norms to produce models of the impact of feature typicality upon written word processing. Their work clearly illustrates the important point that concept typicality is determined by the intercorrelation between component features, rather than their number or frequency: an apple is a more typical fruit than an avocado by virtue of the stronger intercorrelation of its features with those of other fruits. The impact of concept typicality observed here under conditions of semantic impairment is in accordance with this notion that the connections within the semantic layer are determined by their degree of intercorrelation—it is an "S<>S" effect. This is consistent with the finding that concept typicality effects are also seen in semantic judgments upon objects and written words, which do not require spoken output (Morrison et al., 1992; Larochelle and Pineau, 1994; Holmes and Ellis, 2006). Although both AoA and frequency were strong determinants of picture naming performance for the SD patients, neither variable interacted with lATL rTMS, in contrast to the interaction seen for concept typicality. These results introduce the possibility that the impact of AoA and frequency may derive from their influence on connections outside the semantic system, in contrast to Steyvers and Tenenbaum (2005) proposal that AoA influences the formation of semantic representations themselves.

It is possible that the effects of AoA and frequency seen here at least partially derive from the speech production aspect of the picture naming task, which is consistent with the observation of weak effects of these variables relative to typicality in semantic categorization tasks (Barbon and Cuetos, 2006). Indeed, Lambon Ralph and Ehsan (2006) have demonstrated that while AoA may be considered a semantic effect in that it is more influential in picture naming than reading aloud, this is because its influence is most pronounced in tasks involving arbitrary input-output mappings. Given that the model they presented was successfully able to simulate this effect without any within semantic level connections (S<>S), it would seem that AoA may determine the weights on the connections mapping between semantics and phonology ("S>P"). Further, it may well be that the name frequency variable considered here exerted its influence primarily on the connections within the phonological output layer ("P<>P"), although the influence of frequency is pervasive within connectionist models (Plaut et al., 1996). Further simulation work explicitly considering the impact of concept typicality upon naming within connectionist models is clearly required to determine the locus of these effects, which in turn can generate hypotheses about the neural regions in which the impact of different stimulus variables should be most apparent.

The impact of concept typicality under conditions of semantic disruption observed here speaks to the importance of this variable in the representation of meaning, but this conclusion applies of course only to concrete concepts, due to the use of the picture naming task. To date, the vast bulk of research into semantic representation has focused on concrete concepts, which for the most part fall into natural categories, making concept typicality a pertinent dimension. This does, however, seem less applicable to the representation of abstract concepts, which form the majority of our semantic knowledge. Nevertheless, if the dimension of concept typicality is implemented in terms of the degree of intercorrelation of component features, this same general principle would also seem applicable to abstract concepts. Abstract concepts tend to have lower feature intercorrelation than concrete concepts by virtue of their more fluid and contextually dependent meanings (Harm and Seidenberg, 2004). As abstract concepts are more likely to have meanings that vary according to context, only a subset of their features are activated together on any given occasion, and thus their components co-occur less reliably than is the case for concrete concepts. Indeed, recent case-series studies have found abstract concepts to be more vulnerable to degradation than concrete concepts in SD (Jefferies et al., 2009; Hoffman and Lambon Ralph, 2011) and after lATL rTMS (Pobric et al., 2009). The common mechanism of differential feature intercorrelation functions to align these concreteness effects with the concept typicality effects found in the present study.

The unique influence of concept typicality upon picture naming performance under conditions of semantic disruption reported here illustrates the complementary roles that case-series neuropsychology and rTMS play in revealing the nature of the meaning that we have in mind and where it resides in the brain. The greater susceptibility of lower than higher typicality concepts to ATL damage in SD combined with the selective effect of lATL rTMS on naming of lower relative to higher typicality concepts clearly provides target data for future connectionist models of semantic representation and generates hypotheses for neuroimaging studies. This investigation has shown that a key aspect of conceptual representation is the extent to which component semantic features co-occur, with higher typicality items composed of features which are often activated together, leading to richer and more robust semantic representations. Hence although apples are not the only fruit, they are the most typical by virtue of their highly intercorrelated semantic features.

ACKNOWLEDGMENTS

I am very grateful to Professor John Hodges for permission to publish data from patients in his care, to Professor Karalyn Patterson for helpful discussion of the issues considered in this paper, and to Dr. Gaston Madrid for collecting the rTMS data. The rTMS work reported here was funded by a Leverhulme Trust Research Fellowship RF/2/RFG/2008/0298.

REFERENCES

Adlam, A. L. R., Patterson, K., Rogers, T. T., Nestor, P. J., Salmond, C. H., Acosta-Cabronero, J., and Hodges, J. R. (2006). Semantic dementia and fluent primary progressive aphasia: two sides of the same coin? *Brain* 129, 3066–3080.

Baayen, R. H., Piepenbrock, R., and van Rijn, H. (1993). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Barbon, A., and Cuetos, F. (2006). Age-of-acquisition effects in semantic categorization tasks. *Psicologica* 27, 207–223.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., and Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia* 38, 1207–1215.

Caine, D., Breen, N., and Patterson, K. (2009). Emergence and progression of 'non-semantic' deficits in semantic dementia. *Cortex* 45, 483–494.

Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428.

Davis, C. J. (2005). N-watch: a program for deriving neighborhood size and other psycholinguistic statistics. *Behav. Res. Methods* 37, 65–70.

Dilkina, K., McClelland, J. L., and Plaut, D. C. (2008). A single-system account of semantic and lexical deficits in five semantic dementia patients. *Cogn. Neuropsychol.* 25, 136–164.

Dilkina, K., McClelland, J. L., and Plaut, D. C. (2010). Are there mental lexicons? The role of semantics in lexical decision. *Brain Res.* 1365, 66–81.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198.

Forster, K. F., and Forster, J. C. (2003). DMDX: a Windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124.

Gordon, E., Rohrer, J. D., Kim, L. G., Omar, R., Rossor, M. N., Fox, N. C., and Warren, J. D. (2010). Measuring disease progression in frontotemporal lobar degeneration: a clinical and MRI study. *Neurology* 74, 666–673.

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., and Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology* 76, 1006–1014.

Harm, M. W., and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111, 662–720.

Hoffman, P., and Lambon Ralph, M. A. (2011). Reverse concreteness effects are not a typical feature of semantic dementia: evidence for the hub-and-spoke model of conceptual representation. *Cereb. Cortex* 21, 2103–2112.

Holmes, S. J., and Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Vis. Cogn.* 13, 884–910.

Howard, D., and Patterson, K. (1992). *Pyramids and Palmtrees: A Test of Semantic Access from Pictures and Words*. Bury St. Edmunds: Thames Valley Test Company.

Jefferies, E., Patterson, K., Jones, R. W., and Lambon Ralph, M. A. (2009). Comprehension of concrete and abstract words in semantic dementia. *Neuropsychology* 23, 492–499.

Lambon Ralph, M. A., and Ehsan, S. (2006). Age of acquisition effects depend on the mapping between representations and the frequency of occurrence: empirical and computational evidence. *Vis. Cogn.* 13, 928–948.

Lambon Ralph, M. A. L., Graham, K. S., Ellis, A. W., and Hodges, J. R. (1998). Naming in semantic dementia - What matters? *Neuropsychologia* 36, 775–784.

Lambon Ralph, M. A., McClelland, J. L., Patterson, K., Galton, C. J., and Hodges, J. R. (2001). No right to speak? The relationship between object naming and semantic impairment: neuropsychological evidence and a computational model. *J. Cogn. Neurosci.* 13, 341–356.

Lambon Ralph, M. A., and Patterson, K. (2008). Generalization and differentiation in semantic memory. *Ann. N.Y. Acad. Sci.* 1124, 61–76.

Lambon Ralph, M. A., Sage, K., Jones, R. W., and Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2717–2722.

Larochelle, S., and Pineau, H. (1994). Determinants of response times in the semantic verification task. *J. Mem. Lang.* 33, 796–823.

Lezak, M. (1976). *Neuropsychological Assessment*. New York, NY: Oxford University Press.

Loftus, G. R., and Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychon. Bull. Rev.* 1, 476–490.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.

Mayberry, E. J., Sage, K., and Lambon Ralph, M. A. (2011). At the edge of semantic space: the breakdown of coherent concepts in semantic dementia is constrained by typicality and severity but not modality. *J. Cogn. Neurosci.* 23, 2240–2251.

McRae, K., Cree, G. S., Westmacott, R., and De Sa, V. R. (1999). Further evidence for feature correlations in semantic memory. *Can. J. Exp. Psychol.* 53, 360–373.

McRae, K., De Sa, V. R., and Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *J. Exp. Psychol. Gen.* 126, 99–130.

Morrison, C. M., Chappell, T. D., and Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Q. J. Exp. Psychol. A* 50, 528–559.

Morrison, C. M., Ellis, A. W., and Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Mem. Cognit.* 20, 705–714.

Morrow, L. I., and Duffy, M. F. (2005). The representation of ontological category concepts as affected by healthy aging: normative data and theoretical implications. *Behav. Res. Methods* 37, 608–625.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9, 97–113.

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987.

Patterson, K., Ralph, M. A. L., Jefferies, E., Woollams, A., Jones, R., Hodges, J. R., and Rogers, T. T. (2006). "Presemantic" cognition in semantic dementia: six deficits in search of an explanation. *J. Cogn. Neurosci.* 18, 169–183.

Plaut, D. C. (1996). Relearning after damage in connectionist networks: toward a theory of rehabilitation. *Brain Lang.* 52, 25–82.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115.

Pobric, G., Lambon Ralph, M. A., and Jefferies, E. (2009). The role of the anterior temporal lobes in the comprehension of concrete and abstract words: rTMS evidence. *Cortex* 45, 1104–1110.

Pobric, G., Jefferies, E., and Lambon Ralph, M. A. (2010a). Amodal semantic representations depend on both anterior temporal lobes: evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia* 48, 1336–1342.

Pobric, G., Jefferies, E., and Lambon Ralph, M. A. (2010b). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Curr. Biol.* 20, 964–968.

Pobric, G., Jefferies, E., and Ralph, M. A. L. (2007). Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20137–20141.

Raven, J. C. (1962). *Coloured Progressive Matrices: Sets A, AB, B*. London: Lewis.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., and Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol. Rev.* 111, 205–235.

Rogers, T. T., and McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press.

Rohrer, J. D., McNaught, E., Foster, J., Clegg, S. L., Barnes, J., Omar, R., Warrington, E. K., Rossor, M. N., Warren, J. D., and Fox, N. C. (2008). Tracking progression in frontotemporal lobar degeneration: serial MRI in semantic dementia. *Neurology* 71, 1445–1451.

Rohrer, J. D., Warren, J. D., Modat, M., Ridgway, G. R., Douiri, A., Rossor, M. N., Ourselin, S., and Fox, N. C. (2009). Patterns of cortical thinning in the language variants of frontotemporal lobar degeneration. *Neurology* 72, 1562–1569.

Rosch, E., and Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognit. Psychol.* 7, 573–605.

Smith, E. E., Shoben, E. J., and Rips, L. J. (1974). Structure and process in semantic memory: a featural model for semantic decisions. *Psychol. Rev.* 81, 214–241.

Steyvers, M., and Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cogn. Sci.* 29, 41–78.

Visser, M., Embleton, K. V., Jefferies, E., Parker, G. J., and Ralph, M. A. L. (2010). The inferior, anterior temporal lobes and semantic memory clarified: novel evidence from distortion-corrected fMRI. *Neuropsychologia* 48, 1689–1696.

Wechsler, D. A. (1987). *Weschler Memory Scale: Revised*. New York, NY: Psychological Corporation.

Woollams, A. M., Cooper-Pye, E., Hodges, J. R., and Patterson, K. (2008). Anomia: a doubly typical signature of semantic dementia. *Neuropsychologia* 46, 2503–2514.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 January 2012; accepted: 25 March 2012; published online: 18 April 2012.

Citation: Woollams AM (2012) Apples are not the only fruit: the effects of concept typicality on semantic representation in the anterior temporal lobe. Front. Hum. Neurosci. ***6****:85. doi: 10.3389/fnhum.2012.00085*

APPENDIX

Table A1 | Names of pictures used as stimuli.

Low typicality	High typicality
SD PATIENTS	
Accordion	Aeroplane
Alligator	Apple
Camel	Axe
Cherry	Banana
Cooker	Bicycle
Crocodile	Bus
Desk	Cat
Eagle	Chicken
Fox	Cow
Fridge	Deer
Frog	Dog
Guitar	Drum
Harp	Duck
Helicopter	Elephant
Kangaroo	Hammer
Lamp	Horse
Lobster	Lion
Motorbike	Lorry
Mouse	Monkey
Ostrich	Orange
Owl	Pear
Paintbrush	Piano
Peacock	Pliers
Penguin	Rabbit
Pineapple	Saw
Pram	Scissors
Rhinoceros	Screwdriver
Rocking chair	Spanner
Seahorse	Stool
Seal	Strawberry
Sledge	Tiger
Squirrel	Tomato
Swan	Train
Tortoise	Trumpet
Zebra	Violin
lATL rTMS	
Acorn	Apple
Balloon	Bed
Bat	Car
Bath	Carrot
Bell	Cat
Boot	Chain
Camel	Chair
Caravan	Coat
Celery	Cow
Cherry	Dog
Crab	Dress
Crown	Drum
Giraffe	Elephant
Harp	Flute
Kangaroo	Fly
Kite	Fork
Peacock	Guitar
Pencil	Hammer
Pepper	Horse
Pram	Knife
Pumpkin	Lemon
Rocket	Lion
Ruler	Orange
Scissors	Pig
Screw	Rabbit
Seahorse	Ring
Shoe	Screwdriver
Snail	Sheep
Snake	Shirt
Squirrel	Sock
Tie	Spanner
Tomato	Spider
Torch	Spoon
Tortoise	Strawberry
Tractor	Table
Typewriter	Train
Waistcoat	Trousers
Whale	Trumpet
Wheelbarrow	Van
Whistle	Violin

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 04 May 2012
doi: 10.3389/fnhum.2012.00106

A combination of thematic and similarity-based semantic processes confers resistance to deficit following left hemisphere stroke

Solène Kalénine[1,2]*, Daniel Mirman[1] and Laurel J. Buxbaum[1]

[1] *Moss Rehabilitation Research Institute, Philadelphia, PA, USA*
[2] *Laboratoire Unité de REcherche en sciences Cognitives et Affectives, Université Charles-de-Gaulle Lille 3, Lille, France*

Edited by:
Penny M. Pexman, University of Calgary, Canada

Reviewed by:
Ken McRae, University of Western Ontario, Canada
Lara Jones, Wayne State University, USA

****Correspondence:***
Solène Kalénine, Laboratoire Unité de REcherche en sciences Cognitives et Affectives, Université Charles-de-Gaulle Lille 3, Domaine universitaire du Pont de Bois, BP 149, 59653 Villeneuve d'Ascq Cedex, France.
e-mail: solene.kaleni@univ-lille3.fr

Semantic knowledge may be organized in terms of similarity relations based on shared features and/or complementary relations based on co-occurrence in events. Thus, relationships between manipulable objects such as tools may be defined by their functional properties (what the objects are used for) or thematic properties (e.g., what the objects are used with or on). A recent study from our laboratory used eye-tracking to examine incidental activation of semantic relations in a word–picture matching task and found relatively early activation of thematic relations (e.g., broom–dustpan), later activation of general functional relations (e.g., broom–sponge), and an intermediate pattern for specific functional relations (e.g., broom–vacuum cleaner). Combined with other recent studies, these results suggest that there are distinct semantic systems for thematic and similarity-based knowledge and that the "specific function" condition drew on both systems. This predicts that left hemisphere stroke that damages either system (but not both) may spare specific function processing. The present experiment tested these hypotheses using the same experimental paradigm with participants with left hemisphere lesions ($N = 17$). The results revealed that, compared to neurologically intact controls ($N = 12$), stroke participants showed later activation of thematic and general function relations, but activation of specific function relations was spared and was significantly earlier for stroke participants than controls. Across the stroke participants, activation of thematic and general function relations was negatively correlated, further suggesting that damage tended to affect either one semantic system or the other. These results support the distinction between similarity-based and complementarity-based semantic relations and suggest that relations that draw on both systems are relatively more robust to damage.

Keywords: semantic processing, thematic knowledge, functional similarity, eye-tracking, stroke

INTRODUCTION

Growing evidence indicates that several types of semantic relationships between objects inform conceptual structure. Both similarity relations based on shared features (also referred to as taxonomic relations, e.g., hammer–screwdriver) and complementary relations based on co-occurrence in events or situations (also referred to as thematic relations, e.g., hammer–nail, see Estes et al., 2011) influence conceptual processing. The degree of feature overlap between concepts predicts the magnitude of semantic priming and semantic competition effects (e.g., Cree et al., 1999; Vigliocco et al., 2004; Mirman and Magnuson, 2009). Similarly, thematic relationships affect semantic priming and categorization behaviors (e.g., Moss et al., 1995; Lin and Murphy, 2001; Hare et al., 2009). Furthermore, recent data support the idea that similarity-based and thematic knowledge are subserved by two functionally distinct systems (Kalénine et al., 2009; Crutch and Warrington, 2010; Schwartz et al., 2011; Mirman and Graziano, 2012). In healthy adults, taxonomic and thematic relationship processing efficiency differs as a function of individual preferences (Mirman and Graziano, 2012) and object kinds (Kalénine and Bonthoux, 2008; Kalénine et al., 2009). Dissociations between taxonomic and thematic knowledge have also been reported in brain-damaged patients, suggesting that the two systems rely on distinct neuroanatomical substrates (Schwartz et al., 2011). The thematic knowledge system would selectively involve areas of the posterior temporal and parietal cortices (Kalénine et al., 2009; Schwartz et al., 2011). The similarity-based system would recruit areas of the anterior temporal lobes (Schwartz et al., 2011) and possibly cerebral regions associated with perceptual similarity processing in the visual cortex (Kalénine et al., 2009). Although the neural delimitations of the two systems have not been fully identified yet and may depend on stimulus and task characteristics, previous evidence suggests that semantic processing could draw on two functionally and neuroanatomically separate systems based on feature similarity and thematic relation computation.

However, the contribution of the two systems to processing of different semantic relationships is not always clear *a priori*. Many relevant semantic relationships may exist for a single object that do not strictly map onto the taxonomic/thematic distinction. This complexity can be easily illustrated with the multiple semantic relationships associated with manipulable object functional use. There are thematic relations that bind objects that are directly used together (e.g., broom is *used with* dustpan). Their processing has been differentiated from taxonomic (feature-based) processing in explicit categorization tasks (Kalénine and Bonthoux, 2008; Kalénine et al., 2009), although thematically related objects may in certain cases share some functional features (e.g., broom and dustpan are *used for* cleaning the floor). There are also functional similarity relationships between objects that share functional features (e.g., broom and vacuum cleaner are *used for* cleaning the floor), which are not necessarily used together directly but might be, occasionally. Moreover, as evidenced by feature generation studies (e.g., Cree and McRae, 2003; McRae, Cree, et al., 2005), a given object may have several functional features of different generality levels (e.g., *used for* cleaning the floor, *used for* cleaning the house), which could lead to the computation of several functional similarity relationships at different levels of generality (broom and vacuum cleaner are *used for* cleaning the floor; and broom, vacuum cleaner, and sponge are *used for* cleaning the house). In this context, it is difficult to determine whether processing functional similarity and thematic relationships would be systematically dissociable, and what would be the contribution of similarity-based and thematic knowledge system to semantic processing of the different relationships associated with object functional use. One may assume that regardless of the level generality of functional features shared by objects, functional similarity relationships would bear upon feature similarity computation (in the present case, functional feature similarity), and would be therefore equally dissociable from thematic knowledge processing. Recent evidence indicates, however, that processing specific functional similarity relationships is likely to involve both similarity-based and thematic knowledge systems.

Results come from a recent study in healthy adults using the "visual world" paradigm (VWP; Kalénine et al., 2012). In the VWP, a set of pictures with experimentally controlled relationships are presented to a participant, and eye movements are recorded while the participant locates the target given an auditory prompt. A key feature of the VWP is that, prior to target identification, distractor pictures that are related or similar to the target in some way compete for attention and are fixated more compared to unrelated distractor pictures. The relation between target and related distractor can be semantic (Huettig and Altmann, 2005; Yee and Sedivy, 2006; Mirman and Magnuson, 2009), phonological (Allopenna et al., 1998), visual (Dahan and Tanenhaus, 2005), or motor (Myung et al., 2006). For an example in the semantic domain, when participants hear the target word "key" and are presented with a four-picture display including the target object (key), a semantically related distractor (lock), and two unrelated distractors (deer and apple), they look more to the lock than to the unrelated distractors before clicking on the key. This pattern reflects the activation of the information shared by the target and related distractors (keys are used on locks) when identifying the target word (key). VWP has several major advantages. The shape of the competition effect can reveal the precise *temporal dynamics* of conceptual activation, in addition to the *magnitude* of conceptual activation (Allopenna et al., 1998; Mirman and Magnuson, 2009). The task is very simple and highly sensitive, so it can reveal subtle differences in conceptual activation in both directions (e.g., greater *vs.* smaller or earlier *vs.* later competition effect), without facing ceiling or floor effect limitations and without introducing complex task demands. These characteristics make the paradigm optimal for the assessment of semantic processing differences in various populations, including very young children and cognitively impaired participants (Huang and Snedeker, 2009; Myung et al., 2010; Silverman et al., 2010; Mirman et al., 2011).

In our study comparing thematic and functional similarity processing in healthy adults, a target word (e.g., broom) was presented in three conditions: with a thematically related distractor picture (e.g., dustpan) in the *Thematic* condition, with a distractor picture that shares the same specific function (vacuum cleaner, *cleaning the floor*) in the *Specific Function* condition, and with a distractor picture that shares the same general function (sponge, *cleaning the house*) in the *General Function* condition. Results showed a competition effect for each of the three types of related distractors of approximately equal magnitude, indicating that thematic, specific function, and general function relationships were all activated approximately equally when performing a word–picture matching task. However, the time courses of activation differed across the three types of relations. Thematic distractors produced an early transient competition effect, whereas General Function distractors produced a later transient competition effect, suggesting a difference in the time course of activation of thematic and function information. Interestingly, the competition effect in the Specific Function condition exhibited an intermediate pattern: relatively extended competition that started early like the thematic competitors and continued late like the general function competitors. These findings suggest that thematic and general functional relationships rely mostly on somewhat distinct thematic and similarity-based processes, respectively. In contrast, objects sharing a specific function may involve a combination of both processes, causing a mixture of earlier and later activation.

The main goal of the present study was to assess the effect of mild-to-moderate left hemisphere stroke on activation of thematic, specific functional, and general functional relationships. We aimed at identifying patterns of behavioral dissociations in a diverse group of stroke participants that were not selected according to specific lesion location. This approach has been proven to be successful in elucidating different patterns of performance related to differences in neuroanatomic substrate (e.g., Buxbaum et al., 2005; Jax et al., 2006). Its main advantage is avoiding statistically underpowered comparisons of very small groups of participants selected on putative lesion location criteria.

In a sample of diverse individuals with brain-damage, we assumed that, regardless of specific lesion location, a single stroke would be less likely to affect both thematic and similarity-based semantic systems simultaneously than one. Thus, if thematic knowledge *or* functional similarity computation is sufficient to activate specific function relationships, then competition in the Specific Function condition should be more robust to damage in stroke. This account parallels the assumptions of dual-coding theories (Paivio, 1986). For example, dual-coding theories explain the greater robustness of concrete than abstract concepts as a result of concrete concepts' capacity to rely on either linguistic or sensory-motor representations. A somewhat less likely alternative is that specific function relations require both thematic knowledge *and* functional similarity computation, which would predict that competition in the Specific Function condition would be most affected following left hemisphere stroke. A similar hypothesis was formulated to account for evidence that word and face recognition can be impaired separately, but object recognition is impaired when either face or word recognition is impaired (Farah, 1991; but see Buxbaum et al., 1999).

In the VWP described above, we predicted that, compared to a group of neurologically intact participants, left hemisphere stroke participants would show reduced and/or later competition between objects that mostly rely on a single semantic process (Thematic and General Function conditions). A corollary prediction is that there should be a negative correlation between impaired activation of Thematic and General Function relations because individuals will tend to have damage to either one or the other. In contrast, competition should be relatively spared when relationships involve a combination of the two semantic processes (Specific Function condition). A less likely outcome is that Specific Function competition would be most affected by stroke, suggesting that it requires both thematic and functional knowledge to be intact. These predictions were tested in the VWP experiment described below.

MATERIALS AND METHODS

PARTICIPANTS

Seventeen left hemisphere stroke participants (eight females, nine males) took part in the study. Participants were recruited from the Neuro-Cognitive Rehabilitation Research Registry at the Moss Rehabilitation Research Institute (Schwartz et al., 2005) and were at least 6 months post-stroke. Participants over the age of 80 and/or with histories of co-morbid neurologic disorders, alcohol or drug abuse, or psychosis were excluded. The mean age for this group was 57 (SD = 11 years) and mean years of education was 14 (SD = 3). All participants had cortical lesions and showed some phonological, lexical, and/or semantic difficulties as reflected by their scores on the Philadelphia Naming Test (PNT; Roach et al., 1996), the comprehension subtest of the Western Aphasia Battery (WAB-comp: Kertesz, 1982), and the Camel and Cactus Test (CCT; Bozeat et al., 2000). Demographic, lesion, and neuropsychological data are reported in **Table 1**. For comparison, we report data from 12 neurologically intact control subjects selected from Kalénine et al. (2012) such that the control group was matched on age ($M = 63$, SD = 5) and education ($M = 14$, SD = 2) to the group of participants with left hemisphere stroke.

All participants gave informed consent to participate in the behavioral testing in accordance with the guidelines of the IRB of Albert Einstein Healthcare Network, were paid $15/h for their participation, and reimbursed for travel expenses.

STIMULI

Stimuli were 96 color photographs of objects, including 16 reference object pictures, 48 semantically related pictures (16 Thematic, 16 Specific Function, and 16 General Function), and 32 unrelated pictures. All 96 critical pictures had at least 90% name agreement. An additional set of 139 pictures was also used for practice and filler trials. Eight 4-picture displays were derived for each reference object. Three displays were used for critical trials, one in each semantic relationship condition. Three other displays were used for composed filler trials and two served as unrelated filler trials. A complete list of the critical items is provided in **Table A1** in Appendix.

Table 1 | Demographic, neuropsychological, and lesion data from the 17 stroke participants.

Participant	Age (year)	Education (year)	Gender	Handedness	PNT	WABc	CCT	Lesion volume (cm^3)	Approximate lesion location
1	58	13	Male	Right	88.6	85	81	103.9	F, P
2	43	12	Female	Right	77.7	99	94	151.3	T, P
3	51	16	Female	Right	91.4	92	77	51.9	F
4	48	18	Female	Right	55.4	95	55	89.1	T, P
5	53	13	Male	Right	67.4	98.5	81	172.2	F, P
6	67	19	Male	Right	72.0	94	78	84.9	T
7	74	9	Male	Right	51.0	98	81	77.3	F
8	73	20	Male	Right	82.3	88.5	86	41.0	F
9	52	14	Female	Right	66.9	98.5	78	31.4	T, P
10	54	12	Male	Right	50.3	86.5	80	57.6	T, P
11	62	14	Female	Right	86.3	100	88	51.5	P
12	59	15	Male	Right	75.4	89.5	39	195.3	F, T, P
13	61	16	Female	Right	30.3	96	89	73.1	F, T, P
14	67	14	Male	Left	25.1	46	72	67.2	T, P, O
15	33	19	Female	Right	93.1	66	81	63.9	T, P, O
16	68	12	Female	Left	86.3	95	75	15.3	F, P
17	48	14	Male	Right	83.4	85	77	55.7	F

PNT, WABc, and CCT refer to the percentage of correct responses on the Philadelphia Naming Test, the comprehension subtest of the Western Aphasia Battery, and the Camel and Cactus Test. Lesion location: F, frontal; T, temporal; P, parietal, O, occipital.

On critical trials, the reference object (e.g., BROOM) was always the target, one object was related to the target (i.e., the competitor) and the last two objects were semantically and phonologically unrelated to both the target and the competitor. The competitor was thematically related to the target in the Thematic displays (e.g., DUSTPAN; *used with broom*), shared a specific function with the target in the Specific Function displays (e.g., VACUUM CLEANER; *clean the floor*), or shared a general function in the General Function displays (e.g., SPONGE; *clean the house*). Composed filler trials were added to allow the related objects to be targets so that participants would not be able to guess which object was the target based on prior exposure. On those trials, the pictures used for critical trials were rearranged and one of the related pictures became the target. Unrelated filler trials involved novel pictures unrelated to each other, one of them being presented twice as the target.

A large norming procedure was conducted on the stimuli. Results are provided in **Table 2**. Visual and manipulation similarity between the reference objects and their corresponding related and unrelated objects was assessed by asking healthy adults to rate on a 7-point scale to what extent the two object pictures were visually similar and the objects displayed could be manipulated in the same way. Visual similarity ratings were low and equivalent between conditions. Manipulation similarity was slightly higher in the Specific Function relationship condition compared to other conditions. Thus, manipulation similarity ratings were used as a covariate when comparing conditions in the analysis of gaze data.

The type of semantic relatedness between reference and distractor objects was evaluated with in three rating blocks. In the Thematic block, participants had to judge on a 7-point scale to what extent the object on the left (reference object) could be used to act with or upon the object on the right (competitor or unrelated object). In the Function Similarity blocks, participants had to judge "to what extent the two objects are similar if one wants to (specific of general similarity)." For example, they had to evaluate to what extent the broom and vacuum cleaner are similar if ones wants to clean the floor (specific similarity) and if one wants to clean the house (general similarity). The ratings confirmed that related objects in the Thematic relationship condition were consistently used to act with/upon each other ($M = 6.6$). In the same way, related objects in the Specific Function and General Function relationship conditions were judged highly similar in the Specific and General Similarity blocks, respectively ($M = 6.1$ and 5.7). Unrelated objects were not associated with the reference objects in any of the three situations: ratings were very low for the unrelated pairs in the Thematic, Specific Similarity, or General Similarity blocks ($M = 1.5$, 1.25, and 1.35, respectively). Moreover, the data indicated that objects in the Specific Function relationship condition (e.g., broom and vacuum cleaner) were judged equally similar in the Specific and General Similarity blocks ($p = 0.12$), while objects in the General Function relationship condition (e.g., broom and sponge) received systematically higher ratings in the General Similarity block compared to the Specific Similarity block ($p < 0.001$). These data confirmed the hierarchical relation between specific and general functional similarities.

Finally, a corpus-based semantic similarity measure (COALS) was used to assess overall degree of semantic relatedness (Rohde, under review). As clearly demonstrated in the presentation of Rohde et al.'s model, COALS is a measure of semantic similarity based on word co-occurrence computation in large text corpora. The measure reflects the fact that words appearing in similar linguistic contexts convey similar meanings. It accounts for over 70% of the variance in word-pair similarity and synonym judgment tasks – more than HAL, LSA, or WordNet. For this reason, we regard it as a good experiment-external measure of overall semantic similarity. Averaged COALS measures for the word pairs used in this experiment indicate that the related object noun pairs were more semantically similar than unrelated pairs, and the degree of semantic relatedness between the reference object noun and the related object nouns did not significantly differ between conditions. Together with the normative ratings collected, this confirmed that Thematic, Specific Function, and General Function conditions differ in the type of semantic relatedness between targets and competitors, not in the degree or amount of overall semantic relatedness.

Overall, there were $16 \times 8 = 128$ trials, including 48 critical trials: 16 Thematic displays, 16 Specific Function displays, and 16 General Function displays. Ten practice trials were also designed on the same model.

APPARATUS

Gaze position and duration were recorded using an EyeLink 1000 desktop eyetracker at 250 Hz. Stimulus presentation and response recording were conducted by E-Prime software (Psychological Software Tools, Pittsburgh, PA, USA).

PROCEDURE

Participants were seated with their eyes approximately 27″ from a 17″ screen with resolution set to $1{,}024 \times 768$ pixels. Since left hemisphere stroke participants often cannot use their contralesional paretic hand, all participants used their left hand to respond. Participants clicked on a central fixation cross to begin each trial.

Table 2 | Mean values and standard deviations of normative ratings and COALS measures for the thematic, specific function, and general function related and unrelated object pairs.

Semantic relationship	Visual ratings	Manipulation ratings	Thematic ratings	Specific function ratings	General function ratings	COALS measure
Thematic	2.6 (1.5)	2.4 (1.2)	6.6 (0.4)	4.8 (1.2)	5.6 (0.7)	0.17 (0.14)
Specific function	3.4(1.5)	3.9 (1.3)	4.8 (0.9)	6.1 (0.5)	6.4 (0.4)	0.15 (0.14)
General function	2.6 (1.5)	3.0 (1.7)	3.9 (0.8)	3.4 (1.3)	5.7 (0.6)	0.18 (0.16)
Unrelated	2.7 (1.4)	2.1 (0.9)	1.2 (0.3)	1.5 (0.5)	1.3 (0.9)	0.02 (0.04)

Then they saw four images; each image was presented near one of the screen corners. Images had a maximum size of 200 × 200 pixels and were scaled such that at least one dimension was 200 pixels. Therefore, each picture subtended about 3.5˚ of visual angle. The position of the four pictures was randomized. The display was presented for a 1-s preview to allow for initial fixations that are driven by random factors or visual salience rather than word processing. Two hundred and fifty milliseconds before the offset of the preview, a red circle appeared in the center of the screen in order to drive attention back to the neutral central location. Then participants heard the target word through speakers and had to click on the image that corresponded to the target word (**Figure 1**). Eye movements were recorded starting from when the display appeared on the screen and ending when the participant clicked on the target picture. The same procedure was followed for the 10 practice trials and the 128 test trials. The test trial order was randomized.

DATA ANALYSIS

Fixation data averaging

Four areas of interest (AOI) associated with the four object pictures were defined in the display. Each AOI corresponded to a 400 × 300 pixel quadrant situated in one of the four corners of the computer screen. Accordingly, fixations that fell into one of these AOI were considered object fixations, while fixations that fell out of any of the AOI were non-object fixations. At any moment on a single trial, a participant can either fixate an object or not; thus, fixation proportion of each AOI can be either 0 or 1 at any point in time. For each trial of each participant, we computed the proportion of time spent fixating each AOI for each 50 ms time bin. Critical trial data were averaged over items and participants in order to obtain a time course estimate of the fixations on the target, related, and unrelated objects. Data from filler trials were not analyzed. The proportion of fixations on the two unrelated objects was averaged.

Growth curve analysis statistical approach

Growth curve analysis (GCA) is a multi-level modeling framework specifically designed to analyze change over time and adapted for analysis of fixation time course (Magnuson et al., 2007; Mirman et al., 2008). GCA allows simultaneous quantification of fine-grained time course differences between groups and/or conditions of interest as well as between individuals within a group or condition. This is particularly relevant for neuropsychological studies that commonly aim at both comparing a small patient sample to a control group and comparing patients with one another (e.g., Mirman et al., 2008, 2011).

Growth curve analysis of gaze data typically captures the data pattern with two model levels. The first submodel, called *Level-1*, captures the effect of Time on fixation proportions using fourth-order orthogonal polynomials. A fourth-order polynomial is necessary to capture the rise and fall of fixation probabilities over the course of a trial. Specifically, the intercept term reflects average overall fixation proportion, the linear term reflects a monotonic change in fixation proportion (similar to a linear regression of fixation proportion as a function of time), the quadratic term reflects the symmetric rise and fall rate around a central inflection point, and the cubic and quartic terms similarly reflect the steepness of the curve around inflection points. In our paradigm, these higher order terms (i.e., cubic and quartic) appear to distinguish particularly well between early-rising/transient *vs.* later-rising/longer-lasting fixation time courses (Kalénine et al., 2012).

The second set of submodels, called *Level-2*, capture the experimental effects of group, condition, etc. on the Level-1 time terms. They describe each level-1 model term as a function of population means, fixed effects, and random effects. Fixed effects correspond to the effects of the experimental manipulations (group and/or conditions). Random effects can express (a) the deviation for one subject (or item) from the grand mean of fixation proportion (quantification of general individual differences), and (b) the deviation of one subject (or item) in a particular condition from the mean of this participant and the mean of this condition (quantification of individual differences for a particular manipulation). Thus, while fixed effects evaluate the effect of the experimental manipulations at the group level, random effects provide a way to quantify individual participant (or item) effect sizes. Individual effect sizes can then be used to assess individual differences.

Using this multi-level modeling approach, we conducted two separate sets of analyses. First, we compared the patterns of competition for the three Display Types (Thematic, Specific Function, and General Function) within the group of left hemisphere

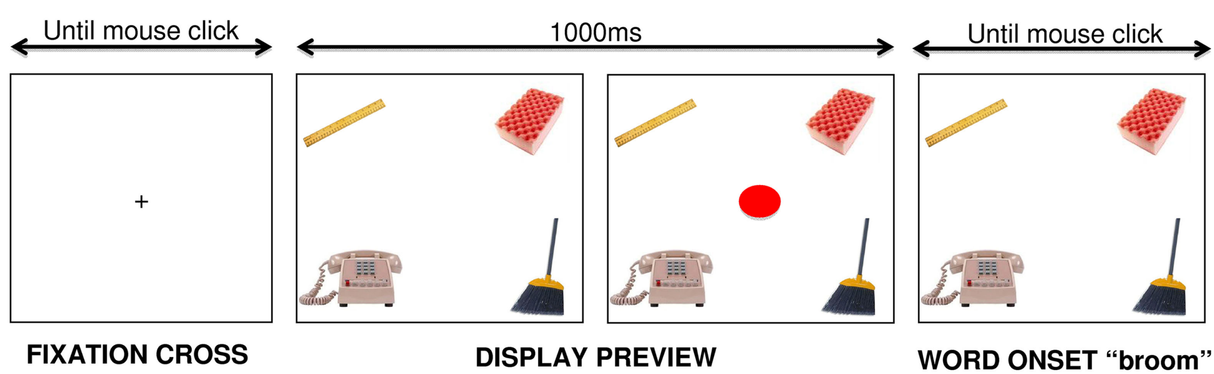


FIGURE 1 | Example of trial used in the eye-tracking experiment. The display presents the target object (e.g., broom), a semantic competitor (e.g., sponge), and two unrelated objects (e.g., phone and ruler). Target words were delivered after a 1000-ms preview of the display.

stroke participants. If, as suggested by the results from our first study (Kalénine et al., 2012), functional similarity and thematic processes are somewhat distinct and the Specific Function pairs draw on a combination of both, we should find (1) differences in the amount and/or time course of competition between General Function and Thematic displays on the one hand, and Specific Function displays on the other hand and (2) a negative correlation between degree of General Function and Thematic competition across stroke participants.

The second analysis compared the two groups – left hemisphere stroke participants and neurologically intact controls – in each of the Display Types. Extending the logic of hypothesis 1, we may find (3) distinct patterns of competition effect differences between stroke and control participants in Thematic and General Function displays on the one hand, and Specific Function displays on the other hand. These predictions are described in more detail below.

Within-group analysis: comparison of the time course of thematic, specific function, and general function competition in stroke participants

In the by-subject analysis, fixation probabilities over time were modeled as a function of Object Relatedness (competitor, unrelated), Display type (Thematic, Specific Function, General Function), and the Object Relatedness × Display Type interaction as fixed effects, with Subject and Subject × Object × Display Type as random effects. In the by-item analysis, the Subject factor was replaced by the Item factor. In addition, since manipulation similarity between objects was known to differ between display types, this factor was introduced as a control variable in the *Level-2* model before the factors of interest in the item analysis.

Fixed effects were incorporated in the *Level-2* submodels incrementally in three (by-subject) or four (by-item) steps. In this way, it was possible to test the improvement of the model fit after adding each factor of interest and, thus, evaluate the overall effects of Object Relatedness, Display Type and the interaction between Object Relatedness and Display Type on the time course of the gaze data, while controlling for differences in manipulation similarity between conditions. Models were fit using Maximum Likelihood Estimation and compared using the −2LL deviance statistic (−2 times the log-likelihood), which is distributed like χ^2 with k degrees of freedom corresponding to the k parameters added.

If activation of Specific Function relations can draw either upon the thematic system or the similarity-based system, then there should be more robust competition in this condition than in the Thematic and General Function displays in stroke participants. In contrast, if both systems are required, then Specific Function competition should be the most impaired condition in stroke participants. As illustrated in Kalénine et al. (2012), we anticipated that the earlier-rise *vs.* later-rise of competition effects should be visible on higher order terms (cubic and/or quartic). Thus, we expected significant competition effect differences between Display Types on these time terms.

Moreover, we used the random effects of this analysis to quantify individual effect sizes in each of the three Display Type conditions. We then examined the correlations of individual participant effects sizes between conditions in order to test the relationships among the competition effect time courses in the Thematic, Specific Function, and General Function conditions (for an example of this approach in the phonological domain, see Mirman et al., 2011). Specifically, the hypothesis that there are distinct thematic and similarity-based processes predicts a negative correlation between individual competition effect sizes in the Thematic and General function displays (because left hemisphere stroke participants will tend to have one kind of damage or the other). In contrast, competition effect sizes in the Specific Function displays should overall not be related to effect sizes in the other conditions.

Between group analysis: comparison of the time course of competition in each display type between stroke and control participants

In the by-subject analysis, fixation probabilities over time were modeled as a function of Object Relatedness (competitor, unrelated), Group (stroke participants, controls), and Object Relatedness × Group as fixed effects, with Subject and Subject × Object as random effects. In the by-item analysis, the Subject factor was replaced by the Item factor. Fixed effects were incorporated incrementally in three *Level-2* submodels. Using the same model comparison approach as in the within-group analysis, we assessed the overall effects of Object Relatedness, Group, and the interaction between Object Relatedness and Group on the time course of fixations, in each condition. We expected an overall effect of group and, more importantly, an interaction between Group and Object Relatedness, which would indicate differences in the competition effect time course between groups. Again, we then tested this interaction on the different time terms.

As described above, the core hypothesis being tested was that Thematic and General Function competition relies mostly on distinct semantic processes, i.e., thematic or feature similarity processing, whereas Specific Function competition draws on both. As a result, we predicted that Thematic and General Function competition would be vulnerable to left hemisphere stroke. Accordingly, we expected stroke participants to show later-rising competition effects compared to controls in the General Function and Thematic displays. In the Specific function displays, if either one process or the other is sufficient to activate the semantic relationship, stroke participants should demonstrate close-to-normal competition effects. Alternatively, if Specific Function relations require both processes, then the Specific Function competition should be later-rising in stroke participants compared to controls. As in the within-group analysis, differences in competition time courses between groups should be particularly obvious on the cubic and/or quartic terms.

RESULTS

All participants, left hemisphere stroke participants and neurologically intact controls, were highly accurate in identifying the target object among distractors in all three conditions, performing on average between 95 and 99% correct (no significant difference between groups or conditions, all $F < 1$). Mean mouse click reaction times from display onset was 3081 ms for the control group and 4536 ms for the stroke participant group [$F(1,78) = 3.90$, $p = 0.052$]. There was no effect of Display Type and no interaction between Group and Display Type on mouse click reaction times ($F < 1$).

Gaze data were collected from the onset of each trial (i.e., the presentation of the four-picture display) to the end of the trial (i.e., the mouse click). No trial had to be excluded because of a lack of gaze data (track loss or off-screen fixations). Each trial received between 2 and 27 fixations in controls ($M = 9$, $SD = 2.6$), and between 1 and 55 fixations in stroke participants ($M = 11.8$, $SD = 4.6$). Trials where participants made an incorrect response or the reaction time was more three standard deviations from the participant's condition mean (1.8% of control data; 3.8% from stroke participant data) were excluded from the fixation analysis.

Figure 2 shows the averaged time course of fixations to the target, competitor and unrelated objects from target word onset for the participants with left hemisphere stroke (top) and for the control participants (bottom). The statistical analysis was restricted to the competition effects driven by the linguistic input. Accordingly, we compared fixation proportion between related and unrelated distractors from 500 ms until 2000 ms after word onset. This analysis window was chosen because it starts slightly before target fixation proportions begin to rise above distractor fixations (i.e., when fixations start to be driven by processing of the target word) and ends when the competition has been resolved and target fixation proportions have reached their ceiling[1].

WITHIN-GROUP ANALYSIS: COMPARISON OF THE TIME COURSE OF THEMATIC, SPECIFIC FUNCTION, AND GENERAL FUNCTION COMPETITION IN LEFT HEMISPHERE STROKE PARTICIPANTS

Overall, there was neither an effect of Object Relatedness [by-subject: $\chi^2(5) = 5.89$, $p = 0.31$; by-item: $\chi^2(5) = 4.90$, $p = 0.42$] nor an effect of Display Type [by-subject: $\chi^2(10) = 12.15$, $p = 0.27$; by-item: $\chi^2(5) = 14.43$, $p = 0.15$]. However, there was a reliable Object Relatedness × Display Type interaction [by-subject: $\chi^2(10) = 24.17$, $p < 0.01$; by-item: $\chi^2(10) = 25.98$, $p < 0.005$] indicating differences in the time course of competition across the three types of competitors.

Significance tests on the individual parameter estimates revealed that there was no difference in the overall amount of competition between the display types (intercept term: all $p > 0.30$).

[1]For a meaningful comparison with the gaze data from stroke participants, the time window used in the prior analysis of the neurologically intact participant data (Kalénine et al., 2012) has been extended from 1300 to 2000 ms after word onset.

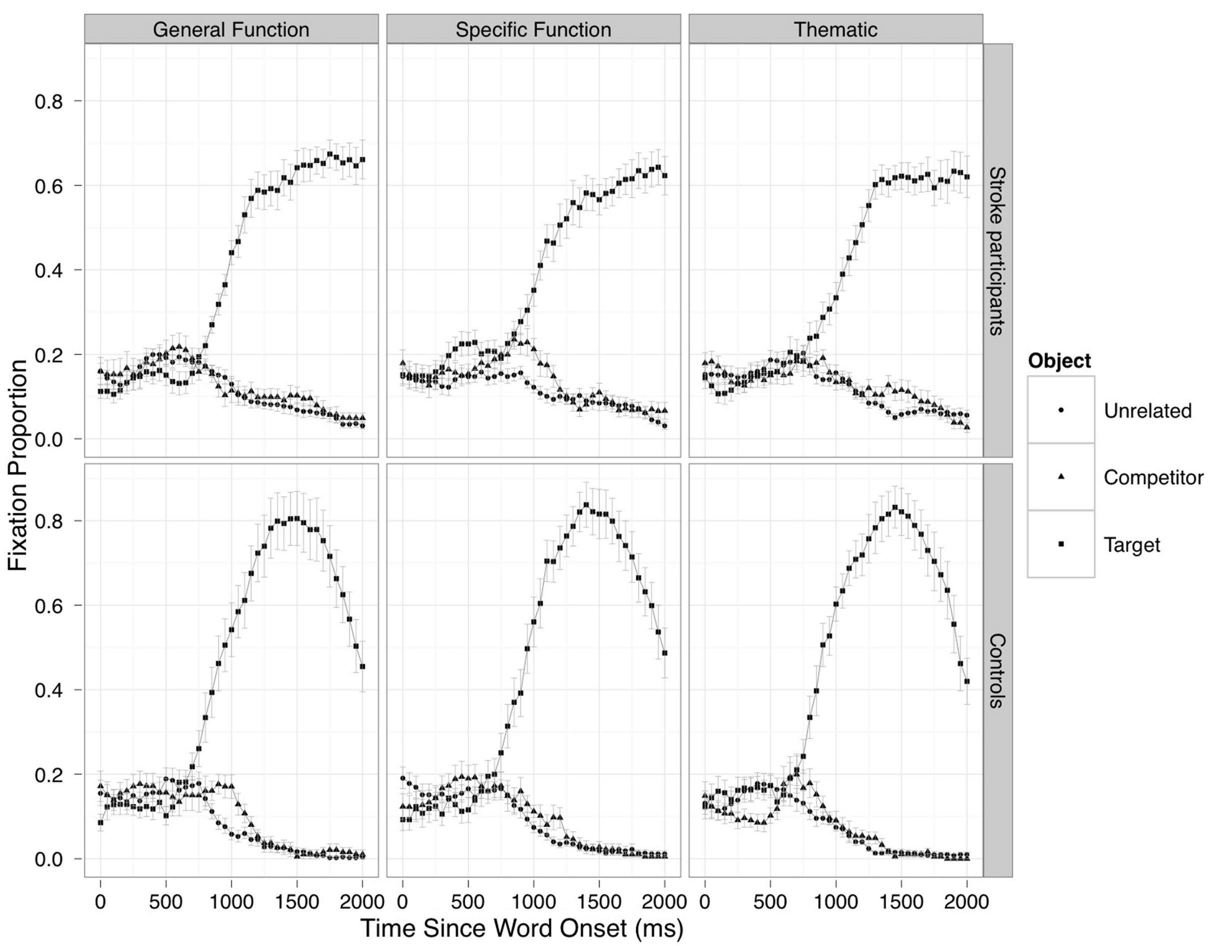


FIGURE 2 | Averaged time course of fixations to the target, competitor and unrelated objects from word onset in each display type for stroke (top) and control (bottom) participants.

However, the Specific Function display significantly differed from the other two display types on the cubic term (Specific Function–General Function: Estimate = 0.182, SE = 0.065, $p < 0.01$ by-subject, and Estimate = 0.173, SE = 0.059, $p < 0.01$ by-item; Specific Function–Thematic: Estimate = 0.173, SE = 0.067, $p < 0.05$ by-subject, and Estimate = 0.164, SE = 0.062, $p < 0.01$ by-item), reflecting the earlier-rising and more transient competition effect in this condition compared to the other two (**Figure 2**, top row).

The correlation analysis between individual competition effects sizes in the three conditions indicated that competition effect time courses in the Thematic and General Function displays were negatively correlated (**Figure 3**). In particular, stroke participants who showed a greater amount and rise of fixations to the competitor in the Thematic condition also tended to have a reduced amount and rise of fixations to the competitor in the General Function (Intercept: $r = -0.50$, $p < 0.05$, **Figure 3A**; Linear: $r = -0.65$, $p < 0.05$, **Figure 3B**). Individual competition effect sizes in the Specific Function condition were not reliably correlated with effect sizes in either of the other conditions.

BETWEEN GROUP ANALYSIS: COMPARISON OF THE TIME COURSE OF COMPETITION IN EACH DISPLAY TYPE BETWEEN LEFT HEMISPHERE STROKE AND CONTROL PARTICIPANTS

In the General Function displays (**Figure 2**, left column), there was an effect of group on the overall time course of fixations, regardless of object relatedness [by-subject: $\chi^2(5) = 16.14$, $p < 0.01$; by-item: $\chi^2(5) = 256.29$, $p < 0.0001$]. The interaction between Group and Object Relatedness failed to reach significance in the by-subject analysis [$\chi^2(5) = 8.76$, $p = 0.11$], but was highly reliable in the by-item analysis [$\chi^2(5) = 41.48$, $p < 0.0001$], suggesting differences in the time course of the competition effect between stroke participants and neurologically intact controls. Significance tests on the parameter estimates showed a later-rising but longer-lasting competition effect for stroke participants compared to controls, as indicated by a reliable difference between groups on the cubic term (Estimate = −0.171, SE = 0.062, $p < 0.01$ by-subject; Estimate = −0.142, SE = 0.027, $p < 0.0001$ by-item).

The same pattern was observed in the Thematic displays (**Figure 2**, right column). There was an overall effect of Group [by-subject: $\chi^2(5) = 22.64$, $p < 0.001$; by-item: $\chi^2(5) = 451.33$, $p < 0.0001$], and an interaction between Group and Object Relatedness, highly significant by-item [by-subject: $\chi^2(5) = 5.50$, $p = 0.35$; by-item: $\chi^2(5) = 60.17$, $p < 0.0001$]. This interaction was clearly visible on the cubic term (Estimate = −0.123, SE = 0.057, $p < 0.05$ by-subject; Estimate = −0.112, SE = 0.027, $p < 0.001$ by-item). As in the General Function displays, the Thematic competition effect was later-rising for stroke participants compared to control participants.

In contrast, stroke participants did not show later competition effects than controls in the Specific Function displays (**Figure 2**, middle column). In this condition, there was a reliable effect of Group [by-subject: $\chi^2(5) = 19.98$,

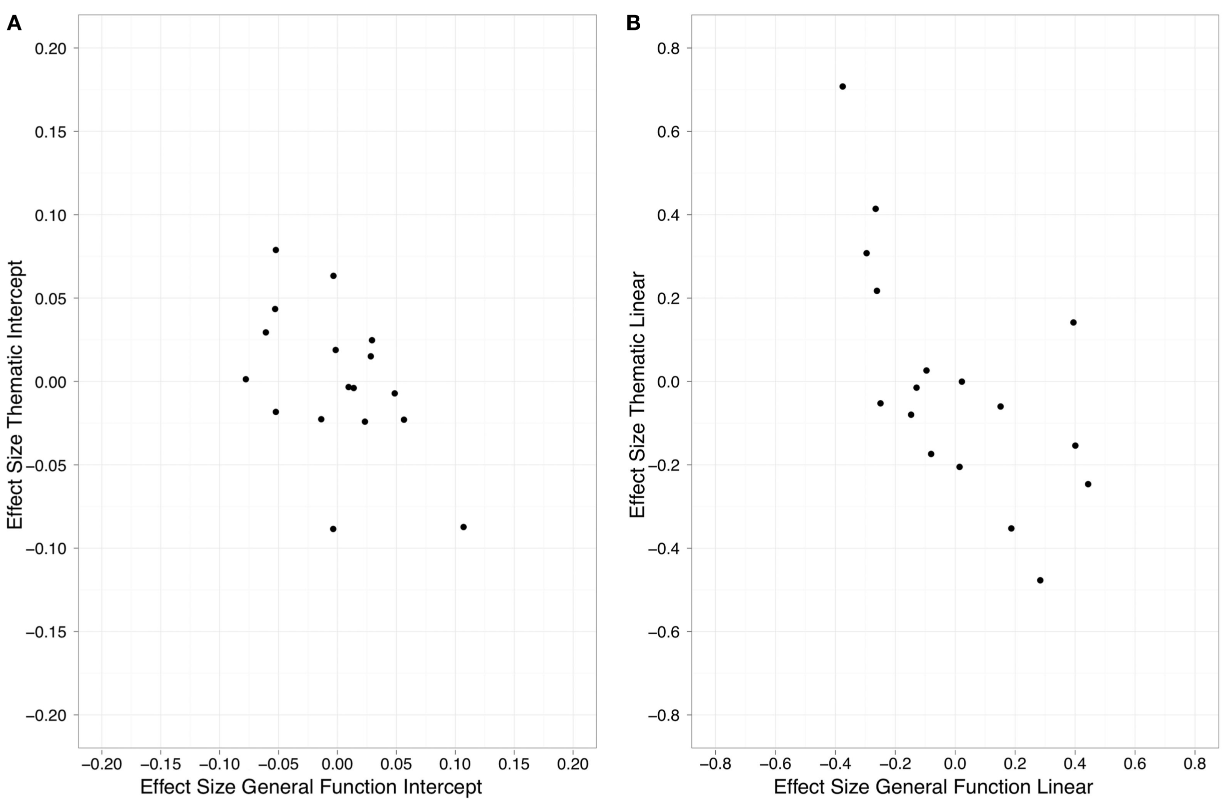


FIGURE 3 | Relationship between individual competition effect estimates in the thematic and general function displays on the intercept (A) and linear (B) terms of the model for the group of stroke participants.

$p < 0.005$; by-item: $\chi^2(5) = 397.64$, $p < 0.0001$], and a significant Group × Object Relatedness interaction by-item [by-subject: $\chi^2(5) = 3.96$, $p = 0.55$; by-item: $\chi^2(5) = 26.33$, $p < 0.0001$]. This interaction tended to be significant on the cubic term (Estimate = 0.110, SE = 0.070, $p = 0.11$ by-subject; Estimate = 0.123, SE = 0.027, $p < 0.0001$ by-item]. Critically, the difference between the competition effect time courses in the two groups was in the opposite direction (positive estimate here, negative estimates in the other display types). That is, the activation of specific function relations tended to be earlier-rising and more transient in stroke participants compared to neurologically intact controls.

DISCUSSION

To sum up, results from the present study showed that (1) Left hemisphere stroke participants exhibited later activation of Thematic and General Function relations than Specific Function relations during the identification of a manipulable artifact object among distractors. (2) Across stroke participants, there was a negative relationship between the competition effects sizes in the Thematic and General Function conditions. (3) Stroke participants exhibited later Thematic and General Function competition effects and earlier Specific Function competition effects compared to a group of age- and education-matched neurologically intact control participants.

We propose that the different temporal dynamics between the three types of semantic relationships reflect the relative involvement of distinct thematic and similarity-based processes in semantic processing of manipulable objects. Processing thematically related objects (e.g., broom–dustpan) mostly relies on thematic knowledge about the roles of objects in events (Nelson, 1983, 1985; McRae, Hare, et al., 2005; Bonthoux and Kalénine, 2007), regardless of object property overlap. In contrast, processing objects related by a general function (e.g., broom–sponge) mainly relies on the computation of the features shared by the two objects. Because these two kinds of knowledge/processing are functionally and neuroanatomically distinct, a given stroke is unlikely to disrupt both processes. Thus, participants with weaker thematic knowledge activation tend to show preserved feature similarity processing, and vice-versa (Schwartz et al., 2011; Mirman and Graziano, 2012; under review).

Recent data from stroke participants suggest that anterior temporal lobe structures are particularly important for taxonomic semantic knowledge and temporo-parietal cortex is particularly important for thematic semantic knowledge (Schwartz et al., 2011; Mirman and Graziano, under review), but in the present study we failed to find any systematic association between percentage damage to these locations and competition effect sizes in the different conditions. It is hard to interpret this null result. It is possible that we did not have enough statistical power to detect this association, but one cannot distinguish between lack of an effect and lack of power. Further studies will be needed to investigate the neuroanatomical bases of thematic and feature similarity processing while considering the various similarity-based relationships a single object may have.

The main novel finding of the present study concerns the relative sparing of specific function relations – the condition that we hypothesized to involve both thematic and functional similarity processes. The reasons for this putative combination in the Specific Function condition in both healthy adults and stroke participants are not clear. One possibility is that the thematic system strongly involves action knowledge processing, especially for manipulable objects. In contrast, we may speculate that the similarity-based system at play in computing functional similarities is less likely to recruit action knowledge. This is consistent with the dissociation observed in certain situations between action and function knowledge (e.g., Buxbaum and Saffran, 2002; Boronat et al., 2005; Canessa et al., 2008; Pelgrims et al., 2011). However, when objects are functionally similar at the specific level, action and function may become more interconnected in a computational sense, which would be reflected by activation of both similarity-based and thematic knowledge systems in processing of specific-level concepts. This interpretation requires further investigation.

More importantly for the present issue, results showed that semantic processing of specific functional similarities is more likely to be preserved after stroke. This argues in favor of the assumption that the two semantic processes are somewhat redundant, and that either can be used in this condition, as in classic dual-coding theories of cognitive processes (e.g., Paivio, 1986). In dual-coding theories, knowledge or processes that are supported by a single code (e.g., linguistic or associative representations of abstract concepts) are more vulnerable to damage than those that are supported by two or more codes (e.g., both linguistic and sensory-motor representations of concrete concepts). In these dual-code situations (e.g., recall a concrete word from memory), one code or the other is sufficient to achieve good performance on the cognitive task. Similarly, the findings reported here suggest that in some situations where both thematic and similarity-based processes are involved, only one or the other is sufficient to ensure object semantic processing. Semantic processing of manipulable objects benefits from the involvement of both thematic and feature similarity processing, which leads to close-to-normal performance in the Specific Function condition in the group of stroke participants. Interestingly, competition between objects related by a specific function was even exaggerated in stroke participants compared to controls. It is tempting to speculate that this may be the result of an impairment in a cognitive process that normally manages competition, a frequently observed deficit following stroke (Gotts and Plaut, 2002; Novick et al., 2005; Jefferies et al., 2008). However, such an account cannot explain why earlier competition is only observed in the Specific Function displays and not the others. We also investigated whether the pattern of competition observed in stroke participants was related to other linguistic/semantic neuropsychological measures (i.e., PNT, CCT, and WAB) and did not find any systematic correlations relationships between individual scores on language and semantic tests and competition effect sizes. Reasons for these earlier Specific Function competition effects, then, remain unclear.

The alteration *vs.* preservation of the efficiency of thematic and functional similarity processing after stroke was evident in the time course of competition effects between semantically related distractors in a word-to-matching task. It was not highlighted in explicit object identification measures (mouse click accuracy or reaction times) or in the *magnitude* of the competition effects between semantically related objects. The ability to detect such subtle abnormalities was made possible by the use of a simple experimental paradigm that is sensitive to time course and a statistical technique well-suited to quantifying group, condition, and individual participant effects. We believe that such methods are particularly useful for the study of fine-grained differences in semantic processes in both cognitively intact and impaired populations.

In conclusion, we have provided evidence supporting a relative involvement of two distinct mechanisms in the processing of semantic relationships between objects. Comparison of the temporal dynamics of conceptual activation between different semantic relationships and between left hemisphere stroke and neurologically intact participants suggests that conditions that rely on both mechanisms are more resistant to brain-damage. Semantic richness may be considered in many ways: in terms of multiplicity of sensori-motor modalities involved in a concept (e.g., Campanella and Shallice, 2011), number of contexts associated (e.g., Yap et al., 2011), density of semantic neighborhoods (e.g., Mirman and Magnuson, 2008; Mirman, 2011), multiplicity of semantic processes at play (e.g., Crutch and Warrington, 2010), etc. The present findings provide additional support to the critical role of semantic richness as a predictive dimension of semantic processing in brain-damaged populations.

ACKNOWLEDGMENTS

This research was supported by NIH R01-NS065049 and James S. McDonnell Foundation #220020190 grant awards to Laurel J. Buxbaum, by NIH grant R01-DC010805 to Daniel Mirman, and by the Moss Rehabilitation Research Institute.

REFERENCES

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439.

Bonthoux, F., and Kalénine, S. (2007). Preschoolers' superordinate taxonomic categorization as a function of individual processing of visual vs. contextual/functional information and object domain. *Cogn. Brain Behav.* 11, 713–731.

Boronat, C. B., Buxbaum, L. J., Coslett, H. B., Tang, K., Saffran, E. M., Kimberg, D. Y., and Detre, J. A. (2005). Distinctions between manipulation and function knowledge of objects: evidence from functional magnetic resonance imaging. *Cogn. Brain Res.* 23, 361–373.

Bozeat, S., Lambon Ralph, M. A., Patterson, K. E., Garrard, P., and Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia* 38, 1207–1215.

Buxbaum, L. J., Glosser, G., and Coslett, H. B. (1999). Impaired face and word recognition without object agnosia. *Neuropsychologia* 37, 41–50.

Buxbaum, L. J., Johnson-Frey, S. H., and Bartlett-Williams, M. (2005). Deficient internal models for planning hand-object interactions in apraxia. *Neuropsychologia* 43, 917–929.

Buxbaum, L. J., and Saffran, E. M. (2002). Knowledge of object manipulation and object function: dissociations in apraxic and nonapraxic subjects. *Brain Lang.* 82, 179–199.

Campanella, F., and Shallice, T. (2011). Manipulability and object recognition: is manipulability a semantic feature? *Exp. Brain Res.* 208, 369–383.

Canessa, N., Borgo, F., Cappa, S. F., Perani, D., Falini, A., Buccino, G., Tettamanti, M., and Shallice, T. (2008). The different neural correlates of action and functional knowledge in semantic memory: an fMRI study. *Cereb. Cortex* 18, 750–751.

Cree, G. S., McRae, K., and McNorgan, C. (1999). An attractor model of lexical conceptual processing: simulating semantic priming. *Cogn. Sci.* 23, 371–414.

Cree, G. S., and McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *J. Exp. Psychol. Gen.* 132, 163–201.

Crutch, S. J., and Warrington, E. K. (2010). The differential dependence of abstract and concrete words upon associative and similarity-based information: complementary semantic interference and facilitation effects. *Cogn. Neuropsychol.* 27, 46–71.

Dahan, D., and Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: conceptually mediated eye movements during spoken-word recognition. *Psychon. Bull. Rev.* 12, 453–459.

Estes, Z., Golonka, S., and Jones, L. L. (2011). Thematic thinking: the apprehension and consequences of thematic relations. *Psychol. Learn. Motiv.* 54, 249–294.

Farah, M. J. (1991). Patterns of co-occurrence among the associative agnosias: implications for visual object representation. *Cogn. Neuropsychol.* 8, 1–19.

Gotts, S. J., and Plaut, D. C. (2002). The impact of synaptic depression following brain damage: a connectionist account of "access/refractory" and "degraded-store" semantic impairments. *Cogn. Affect. Behav. Neurosci.* 2, 187–213.

Hare, M., Jones, M., Thomson, C., Kelly, S., and McRae, K. (2009). Activating event knowledge. *Cognition* 111, 151–167.

Huang, Y. T., and Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: evidence from real-time spoken language comprehension. *Dev. Psychol.* 45, 1723–1739.

Huettig, F., and Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition* 96, 23–32.

Jax, S. A., Buxbaum, L. J., and Moll, A. D. (2006). Deficits in movement planning and intrinsic control in ideomotor apraxia. *J. Cogn. Neurosci.* 18, 2063–2076.

Jefferies, E., Patterson, K., and Lambon Ralph, M. A. L. (2008). Deficits of knowledge versus executive control in semantic cognition: insights from cued naming. *Neuropsychologia* 46, 649–658.

Kalénine, S., and Bonthoux, F. (2008). Object manipulability affects children and adults' conceptual processing. *Psychon. Bull. Rev.* 15, 667–672.

Kalénine, S., Mirman, D., Middleton, E. L., and Buxbaum, L. J. (2012). Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *J. Exp. Psychol. Learn. Mem. Cogn.* doi: 10.1037/a0027626. [Epub ahead of print].

Kalénine, S., Peyrin, C., Pichat, C., Segebarth, C., Bonthoux, F., and Baciu, M. (2009). The sensory-motor specificity of taxonomic and thematic conceptual relations: a behavioral and fMRI study. *Neuroimage* 44, 1152–1162.

Kertesz, A. (1982). *Western Aphasia Battery*. New York: Grune & Stratton.

Lin, E. L., and Murphy, G. L. (2001). Thematic relations in adults' concepts. *J. Exp. Psychol. Gen.* 130, 3–28.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., and Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cogn. Sci.* 31, 133–156.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods* 37, 547–559.

McRae, K., Hare, M., Elman, J. L., and Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Mem. Cognit.* 33, 1174–1184.

Mirman, D. (2011). Effects of near and distant semantic neighbors on word production. *Cogn. Affect. Behav. Neurosci.* 11, 32–43.

Mirman, D., Dixon, J. A., and Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: growth curves and individual differences. *J. Mem. Lang.* 59, 475–494.

Mirman, D., and Graziano, K. M. (2012). Individual differences in the strength of taxonomic versus thematic relations. *J. Exp. Psychol. Gen.* doi: 10.1037/a0026451

Mirman, D., and Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: processing is slowed by near neighbors and speeded by distant neighbors. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 65–79.

Mirman, D., and Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Mem. Cogn.* 37, 1026–1039.

Mirman, D., Yee, E., Blumstein, S. E., and Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: evidence from eye-tracking and computational modeling. *Brain Lang.* 117, 53–68.

Moss, H. E., Ostrin, R. K., Tyler, L. K., and Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: evidence from priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 863–883.

Myung, J.-Y., Blumstein, S. E., and Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition* 98, 223–243.

Myung, J.-Y., Blumstein, S. E., Yee, E., Sedivy, J. C., Thompson-Schill, S. L., and Buxbaum, L. J. (2010). Impaired access to manipulation features in apraxia: evidence from eyetracking and semantic judgment tasks. *Brain Lang.* 112, 101–112.

Nelson, K. (1983). "The derivation of concepts and categories from event representations," in *New Trends in Conceptual Representation: Challenges to Piaget's Theory?*, ed. E. K. Scholnick (Hillsdale, NJ: Erlbaum), 129–149.

Nelson, K. (1985). *Making Sense. The Acquisition of Shared Meaning*. New York: Academic Press.

Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2005). Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cogn. Affect. Behav. Neurosci.* 5, 263–281.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford: Oxford University Press.

Pelgrims, B., Olivier, E., and Andres M. (2011). Dissociation between manipulation and conceptual knowledge of object use in the supramarginalis gyrus. *Hum. Brain Mapp.* 32, 1802–1810.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., and Brecher, A. R. (1996). The Philadelphia Naming Test: scoring and rationale. *Clin. Aphasiol.* 24, 121–133.

Schwartz, M. F., Brecher, A. R., Whyte, J., and Klein, M. G. (2005). A patient registry for cognitive rehabilitation research: a strategy for balancing patients' privacy rights with researchers' need for access. *Arch. Phys. Med. Rehabil.* 86, 1807–1814.

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A. R., Faseyitan, O. K., Dell, G. S., Mirman, D., and Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8520–8524.

Silverman, L. B., Bennetto, L., Campana, E., and Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition* 115, 380–393.

Vigliocco, G., Vinson, D. P., Lewis, W., and Garrett, M. F. (2004). Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. *Cogn. Psychol.* 48, 422–488.

Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.

Yee, E., and Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *J. Exp. Psychol. Learn Mem. Cogn.* 32, 1–14.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 December 2011; accepted: 10 April 2012; published online: 04 May 2012.

Citation: Kalénine S, Mirman D and Buxbaum LJ (2012) A combination of thematic and similarity-based semantic processes confers resistance to deficit following left hemisphere stroke. Front. Hum. Neurosci. ***6****:106. doi: 10.3389/fnhum.2012.00106*

APPENDIX

Table A1 | List of critical items in the Thematic, Specific Function and General Function conditions, and the corresponding functions evaluated in the norms.

Reference object	Thematic related object	Specific function related object	General function related object	Specific function evaluated	General function evaluated
Bat	Baseball	Glove	(Football) Helmet	Playing baseball	Playing sport
Broom	Dustpan	Vacuum cleaner	Sponge	Cleaning floor	Cleaning house
Clippers	Branch	Hedge trimmer	Rake	Cutting branches	Doing yard work
Eraser	Form	White out	Highlighter	Erasing marks	Working on document
Hammer	Nail	Screwdriver	Pliers	Hanging a picture	Fixing the house
Hook	Fish	Net	Fishing hat	Catching fish	Going on fishing trip
Peeler	Carrot	Knife	Can opener	Peeling vegetables	Cooking dinner
Razor	Shaving cream	Tweezers	Toothbrush	Removing hair	Getting ready in the morning
Saw	Wood	Axe	Drill	Cutting wood	Building things
Scissors	Nails	(Nail) Clippers	Lipstick	Giving herself a manicure	Getting ready for a date
Soap	(Bath) Sponge	Shampoo	Toothpaste	Taking a shower	Keeping a good hygiene
Stapler	Papers	Paperclip	Folder	Binding papers together	Organizing documents
Tape	Package	String	Stamp	Wrapping a package	Sending a package
Toaster	Bread	Waffle-iron	Coffee maker	Cooking breakfast food	Preparing breakfast
Whisk	Eggs	Blender	(Grilling) Spatula	Mixing ingredients	Cooking
Zipper	Jeans	Button	Spool	Fixing pants	Sewing

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 11 July 2012
doi: 10.3389/fnhum.2012.00205

Different influences on lexical priming for integrative, thematic, and taxonomic relations

Lara L. Jones[1]* and Sabrina Golonka[2]

[1] *Department of Psychology, Wayne State University, Detroit, MI, USA*
[2] *Department of Psychology, Leeds Metropolitan University, Leeds, UK*

Edited by:
Melvin Yap, National University of Singapore, Singapore

Reviewed by:
Ken McRae, University of Western Ontario, Canada
Chris F. Westbury, University of Alberta Edmonton, Canada
Jen Coane, Colby College, USA

****Correspondence:***
Lara L. Jones, Department of Psychology, Wayne State University, 5057 Woodward Ave., 7th floor, Detroit, MI 48202, USA.
e-mail: larajones@wayne.edu

Word pairs may be integrative (i.e., combination of two concepts into one meaningful entity; e.g., *fruit—cake*), thematically related (i.e., connected in time and place; e.g., *party—cake*), and/or taxonomically related (i.e., shared features and category co-members; e.g., *muffin—cake*). Using participant ratings and computational measures, we demonstrated distinct patterns across measures of similarity and co-occurrence, and familiarity for each relational construct in two different item sets. In a standard lexical decision task (LDT) with various delays between prime and target presentation (SOAs), target RTs and priming magnitudes were consistent across the three relations for both item sets. However, across the SOAs, there were distinct patterns among the three relations on some of the underlying measures influencing target word recognition (LSA, Google, and BEAGLE). These distinct patterns suggest different mechanisms of lexical priming and further demonstrate that integrative relations are distinct from thematic and taxonomic relations.

Keywords: semantic priming, taxonomic, thematic, integrative, relational representation

Lexical priming refers to faster word recognition latencies following the prior or simultaneous presentation of a meaningfully related prime word. For example, *night* would be recognized more quickly as a real word in the English language following *day, moon, dark, evening, summer,* or the indirectly related *sun*. Semantic richness refers to the variability in the information associated with a word's meaning that can facilitate lexical priming of the target following a related prime (Yap et al., 2011). There are several facets of semantic richness that include characteristics of each individual concept within a prime-target pair (i.e., item measures; e.g., frequency, length, imageability, number of senses, number of associates) as well as pair measures reflecting the relation between the pair (e.g., similarity, co-occurrence, word pair frequency). Our purpose of the current research was to demonstrate a distinction across integrative, thematic, and taxonomic relations on these pair measures. Related to this first goal, we also investigated which of these measures were related to target word recognition latencies in a lexical decision task (LDT) within each of the three relations.

RELATIONAL TAXONOMIES AND DEFINITIONS

The first step in investigating the role of relation types in lexical priming is to define, exemplify, and further establish the underlying item dimensions for each relation type. Recent relational taxonomies (Wu and Barsalou, 2009; Santos et al., 2011) include all three types of relations we will focus on in this paper—integrative, thematic, and taxonomic. Integrative relations are inferred during the process of combining two concepts into a plausible subclass of the second concept (Estes and Jones, 2006, 2009; Jones et al., 2008; *wool socks* are socks made of wool; *summer holiday* is a holiday occurring during the summer months). Integrative relations are included among the "forward phrasal associates" prime-target pairs in the Semantic Priming Project (SPP) (Hutchison et al., 2012), which is a readily available large scale study that includes various item and participant factors in addition to lexical decision and naming latencies (for review see Balota et al., 2012). They are denoted in Santos and colleagues taxonomy as "compound continuation forward." Within what McRae and colleagues (2012) describe as the "entity" relation type, integrative relations include the internal component (e.g., *cherry pit*) and external component (e.g., *tricycle pedals*) subtype relations. Notably, earlier relational taxonomies further subdivided such integrative relations into a small and finite number of general relations (e.g., *have, for, in*; Levi, 1978), though others criticized these general relations as being overly vague (Downing, 1977; Estes and Jones, 2006).

Integrative relations have been studied more extensively in conceptual combination studies using relational priming (e.g., Gagné, 2002; Gagné and Shoben, 2002; Estes, 2003b; Gagné and Spalding, 2004, 2009; Estes and Jones, 2006; Spalding and Gagné, 2011) and memory (Jones et al., 2008; Badham et al., 2012) paradigms. Our focus within this paper is on lexical priming, in which the ability to combine the modifier or prime concept with the head noun or target concept into a plausible entity facilitates word recognition of the target word (Estes and Jones, 2009; Badham et al., 2012). As in the prior conceptual combination studies, the activation of a relation between the two concepts also underlies integrative priming.

Thematic relations refer to the link between concepts that occur together in time and space. Thematically related concepts play complementary roles in a given action or event (e.g., *needle—thread; coffee—juice*; Lin and Murphy, 2001; for review see Estes et al., 2011). The "script" relation in the SPP (Hutchison et al., 2012) includes pairs related to a common event (e.g., *rooster—farm*). They are classified by Santos et al. (2011) as an "aspect of an object or situation" and are often denoted as "event" or "situation" or "script" relations (Moss et al., 1995; Chwilla and Kolk, 2005; Hare et al., 2009; Hutchison et al., 2012; McRae et al., 2012; Metusalem et al., 2012). In turn, these event relations include object-location (e.g., *barn—hay*), and person-location (e.g., *hospital—doctor*) relations among other subtypes (Hare et al., 2009).

Taxonomic relations refer to items associated with a category and may be further divided into superordinate (category—exemplar; e.g., *animal—dog*), coordinate (two exemplars of the same category, e.g., *dog—cat*), and subordinate (e.g., *dog—beagle*). Within this study, we limit our taxonomic items to the category co-member or coordinate relations, which are denoted in the SPP as "category" relations (e.g., *cougar—lion*; Hutchison et al., 2012).

Note that these relation types are not mutually exclusive. Indeed there is much overlap with concept pairs often representing two of the three or even all three relations (e.g., *ice-cream—cake*). Integrative and thematic relations may overlap, particularly for the locative subtype of relation. For example, the concepts *hospital* and *doctor* can be integrated to denote a subclass of doctors that work in a hospital and are thematically related in that hospitals and doctors play complementary roles in a given event or situation. However, there are many other pairs that are thematic but not so integrative (e.g., *prescription—doctor*) or that are integrative but not necessarily thematic (e.g., *animal—doctor*). Integrative and taxonomic relations may overlap depending on the similarity between the concepts and, to a lesser degree, on the extent to which the concepts belong to the same *specific* category. Highly similar items that belong to a specific (or sub-) category are less likely to be integrated than less similar ones (Wisniewski, 1997; Costello and Keane, 2000; Estes, 2003a). For example, *cake* and *pie* have the same shape and both belong to the more general "food" category as well as a more specific "dessert food" category. The high similarity between these items makes them difficult to integrate. Other, less similar, items that belong to the same subcategory (e.g., *cake* and *ice-cream*) may also be considered as thematic in that they may play complementary roles in a given scenario or event (*ice-cream* and *cake* may be served together at a party). More typically though, pairs having both a thematic and taxonomic relation will be co-members of a broader category (e.g., *cake* and *coffee*; *wine* and *cheese*; *meat* and *potatoes*; "foods" or "things that can be consumed").

IMPORTANCE OF RELATION TYPE ON LEXICAL PRIMING

Many lexical priming studies have focused on the role of word association and/or feature similarity in lexical priming (Shelton and Martin, 1992; McRae and Boisvert, 1998; Thompson-Schill et al., 1998; Estes and Jones, 2009; Jones, 2010, 2012; in preparation; for review see Lucas, 2000; Hutchison, 2003; Jones and Estes, 2012). Association strength refers to the proportion of a sample in a free association task indicating a particular concept in response to a cue. For example, nearly 82% of participants in the University of South Florida Free Association norms produced *night* for the cue *day*; Nelson et al., 1998). Associations vary in strength with those having no more than 10% of a sample producing a given target considered as only weakly associated and those with more than 20% considered as strongly associated based on Hutchison's (2003) criteria. Word association strengths influence both the magnitude and even the mere presence of lexical priming (Jones, 2010, 2012; in preparation; for review see Moss et al., 1995; Nation and Snowling, 1999; Lucas, 2000; Hutchison, 2003). Therefore, word association strength must be examined as a factor, minimized, and/or equated when examining the influence of relation types on lexical priming. McRae et al. (2012) argued that equating word association strength by eliminating the most strongly associated items from the stimuli set is not an ideal solution because these items represent the best examples of a given relation. However, we chose to include only "pure" (weakly associated) prime-target pairs in the current research in order to better focus on our other variables of interest (e.g., co-occurrence, similarity), which are often related to association strength (Jones, in preparation).

In contrast to the plethora of studies examining the role of association strength, there have been far fewer studies conducted to "distinguish among types of semantic relations" in lexical priming (McRae and Boisvert, 1998, p. 568; see also McRae et al., 2012). So then further research on relations in lexical priming would fill a long-standing gap in the lexical priming literature. Such investigation is important for several reasons. First, it has implications for the development of semantic memory, which is characterized by a conceptual shift from primarily thematic, functional, or instrumental relations in young children (age < 6) to the addition of categorical (taxonomic) relations along with thematic ones beginning around age 7 (Perraudin and Mounoud, 2009; Jones and Estes, 2012; for review see Estes et al., 2011). Moreover, at least two of these relations—taxonomic and thematic—are neuro-anatomically dissociable (Sachs et al., 2008; Mirman et al., 2011; Schwartz et al., 2011). For instance, individuals with acquired language impairments resulting from brain injury or disease often exhibit specific difficulties with some relations but not others (e.g., Schwartz et al., 2011). Likewise, these relations are also expected to exhibit distinct patterns across item measures that have been found to predict lexical priming (e.g., co-occurrence, word pair frequency, similarity). These underlying measures may differentially predict lexical priming across these three relations, which would have important implications for the semantic priming models (e.g., perceptual simulation, compound cue, expectancy generation) that could account for priming effects.

In addition to distinct patterns of underlying correlates in lexical priming, there may also be differences in the magnitude of priming across relations at various SOAs. Prior studies have found evidence of more robust priming effects for thematic than taxonomic items at short SOAs (Sachs et al., 2008; Sass et al., 2009). Using a standard LDT with a short 200 ms SOA, Sachs et al. (2008) found more robust lexical priming effects (PEs; unrelated—related) for thematically related pairs (e.g., *car—garage*; PE = 57 ms) than for taxonomically related pairs (e.g., *car—bus*; PE = 39 ms), and attributed this result to a greater "salience" for the thematically related items. However, these studies did not control for word association strength, which has been shown to produce more robust lexical priming effects particularly at shorter and longer SOAs (<300 ms and >1500 ms; Moss et al., 1995; Jones, 2012; in preparation). So then the greater salience for the thematic pairs may have actually reflected stronger word associations for the thematic than for the taxonomic pairs.

To the extent that integrative, thematic, and taxonomic relations are conceptually distinct, they should exhibit distinctive patterns across pair measures of semantic richness (e.g., co-occurrence, similarity, word pair frequency). Indeed, Maki and Buchanan (2008) found a three-factor structure across 13 underlying variables (LSA, FAS, etc.) in terms of associative, semantic, and thematic knowledge. In turn, these different types of knowledge have been found to differentially influence lexical priming in prior studies (e.g., Chwilla and Kolk, 2005; Jones and Mewhort, 2007; Hare et al., 2009). Using two different sets of items, we examine the extent to which these three relations have distinct patterns on these pair measures of semantic richness and the extent to which these underlying measures differentially predict lexical priming. Studies 1 and 2 consisted of integrative, thematic, and taxonomic prime-target pairs taken from a large-scale study (with different targets within the three relations; e.g., *tuna sandwich, patient nurse, chalk crayon*). Studies 3 and 4 consisted of a smaller set of prime-target pairs with the target held constant among the three relations (*tomato soup, bowl soup, chili soup*). For both item sets, we minimized and equated association strength and assessed local co-occurrence or word pair frequency (Google hits), and global co-occurrence (LSA cosines).

OVERVIEW OF STUDIES 1 AND 2

In Study 1, we assess the extent to which items taken from the SPP differed on two measures of global and local co-occurrence (described further in the subsequent sections) across our three relations. In Study 2, we sought to examine whether target RTs and priming magnitudes would differ across these three relations using the LDT target RTs taken from the 200 ms and 1200 ms SOAs in SPP.

CO-OCCURRENCE

Co-occurrence between primes and targets influence lexical priming. According to compound-cue theory (Ratcliff and McKoon, 1988; McKoon and Ratcliff, 1992), faster RTs for related primes and targets are produced by the joining of prime and target to form a compound cue which is then matched against items in long-term memory. The degree of facilitation for these target RTs is based on the extent to which the prime and target are associated in memory. Co-occurrence can be assessed at varying levels. Local co-occurrence refers to the extent to which the exact prime-target word pair (e.g., *instruction book*) appears in long-term memory, whereas global co-occurrence refers to the co-occurrence of the prime and target within a given text. In the current study, we assess local co-occurrence by the frequency of the word pair in Google and global co-occurrence using LSA cosines. In addition to influencing lexical priming, the extent and type of co-occurrence is predicted to vary among the three relations.

LSA cosines

Latent Semantic Analysis (LSA) is a statistical approach to language learning that is able to capture subtle semantic relationships between words even though it has no knowledge of word meaning or syntax (Landauer and Dumais, 1997). The logic of the approach is that the "psychological similarity between any two words is reflected in the way they co-occur in small sub-samples of language" (Landauer and Dumais, 1997, p. 215). LSA can be applied at a number of levels—for instance, it can be used to compare texts just as well as it can be used to compare words. In general terms, LSA represents words in terms of their occurrence in particular texts. Singular value decomposition and dimension reduction filter the word vectors so that words occurring in similar or same contexts are represented similarly (Kwantes, 2005). The correlation between vectors is given by the cosine, which is a convenient proxy for the similarity between two words. LSA has successfully modeled a number of behaviors related to cognition and language use. For example, Landauer and Dumais (1997) used LSA both to model the typical vocabulary growth rate of school children and to model semantic priming effects. LSA is also able to recognize words that have the same or similar meanings (Landauer and Dumais, 1994). This reflects the multidimensional use of LSA in prior studies as a measure of similarity (Howard and Kahana, 2002; Gagné et al., 2005) and as a measure of more global co-occurrence (Estes and Jones, 2009; Jones, 2010, 2012). Yet Simmons [Golonka] and Estes (2006) found that LSA cosines were only moderately related to similarity ratings of word pairs ($r = 0.36$). Moreover, in an exploratory factor analysis, Maki and Buchanan (2008) found that LSA along with BEAGLE loaded on the text-based factor rather than the similarity factor. So LSA is likely a better measure of co-occurrence than a proxy for similarity.

Google hits

In contrast to LSA cosines, Google hits assess the local co-occurrence or word pair frequencies of the prime-target in informal written language, taking word order into account when the pair is entered in quotes in the search box. For example, "tomato soup" has a much higher number of Google hits than "soup tomato," whereas the LSA cosines are identical for both word orders. In conceptual combination studies (Wisniewski and Murphy, 2005; Murphy and Wisniewski, 2006) and lexical priming studies (Estes and Jones, 2009; Jones, 2010, 2012), Google hits provided a measure of word pair frequency in everyday written language that was moderately correlated with familiarity ratings (rs $= 0.50$ and 0.60, Wisniewski and Murphy, 2005). Moreover, Google hits are often a better measure of local co-occurrence than familiarity ratings, which tend to be restricted in range and more variable across samples. However, this extensive variation in the number of Google hits can be problematic in that the variability may be much greater within one relation than within another. Hence, logarithmic transformed Google hits (henceforth, logGoogle) may be used to compare across different relation types (Estes and Jones, 2009; Jones, 2012). Study 1 was conducted to investigate the differences among integrative, thematic, and taxonomic relations on these measures of co-occurrence.

STUDY 1

One critical difference between the integrative and the other two relations is that by definition the two concepts in integrative relations can combine into a plausible entity that denotes a subtype of the second concept (e.g., an *herb garden, rose garden*, and *vegetable garden* each denote a specific and plausible subtype of garden). Though some thematic items can be combined into a plausible entity (e.g., *playground slide, giraffe zoo*), the combined entity does not as effectively denote a subtype of the second concept (i.e., most playgrounds have slides; most zoos have giraffes). Thus, word pair frequencies (logGoogle hits) should be higher for the integrative pairs than for the thematic and taxonomic pairs. In contrast, both thematic and taxonomic pairs tend to have greater global or textual co-occurrence than the integrative items, due to the complementary roles the concepts share in a given event for the thematic items and the inclusion within the same category and high semantic similarity for the taxonomic items. Hence, global co-occurrence (LSA cosines) should be greater for the thematic and taxonomic pairs than for the integrative pairs.

MATERIALS

The SPP, (Hutchison et al., 2012) consists of 1661 targets selected from the Nelson et al. (1998) norms with the primary associate and a randomly selected other associate paired with each target. Primes and targets were randomly re-paired in the SPP to create unrelated items within each association group. The SPP includes extensive norms taken from the English Lexicon Project (ELP; Balota et al., 2007; http://elexicon.wustl.edu/) as well as target RTs and priming magnitudes from a LDT with a 200 ms SOA and a 1200 ms SOA. To investigate lexical priming across integrative, thematic, and taxonomic relations for only weakly associated items, we selected items having the following relations from the "Other Associates" tab in SPP: forward phrasal associates, script, and category. Next we eliminated all pairs having forward association strengths (FAS) greater than 0.10 so that only weakly associated items would be included. Results of a One-Way ANOVA confirmed equivalent and weak (all Ms < 0.05) FAS, $F < 1$, $p = 0.63$, and backward association strengths, $F < 1$, $p = 0.83$, across the three relations. Then we limited our items to only noun–noun prime-target pairs and removed any items having proper names for the prime or target (e.g., *hawaii hula, christmas santa*) and morphemic repetition between prime and target (e.g., *bank banker*). The final set of items used in Studies 1 and 2 consisted of 89 integrative items, 78 thematic items, and 85 taxonomic items as shown in Appendix A.

RESULTS AND DISCUSSION

We compared the word pair frequencies (logGoogle hits) and the global/textual co-occurrence (LSA cosines) among the three relations using a One-Way ANOVA with Tukey HSD *post-hoc* tests. Results indicated reliable and robust differences among the three relations for both word pair frequencies (logGoogle hits), $F_{(2,\ 249)} = 64.63$, $p < 0.001$, and global co-occurrence (LSA cosines), $F_{(2,\ 249)} = 13.23$, $p < 0.001$. As shown in **Table 1**, logGoogle was highest for the integrative items, $p < 0.001$, followed by the taxonomic items, which were in turn higher than the thematic items, $p < 0.01$. In contrast, the integrative pairs had reliably lower LSA cosines than the thematic and taxonomic pairs (ps < 0.01), which did not differ ($p = 0.29$). In sum, these results demonstrate distinct patterns of co-occurrence for the integrative items (namely, higher word pair frequencies but lower global/textual co-occurrence) in comparison to the thematic and taxonomic relations.

STUDY 2

The purpose of Study 2 was to determine whether the response times and priming effects would differ among the three relations. Recall that Sachs et al. (2008) found more robust priming for associated thematic pairs (*car—garage*) than for their associated taxonomic pairs (*car—bus*) in a standard LDT with a 200 ms SOA. Here we investigate whether such a difference would occur for our weakly associated thematic, taxonomic, and integrative items by comparing the RTs and priming effects (PEs) found in the 200 and 1200 ms SOAs of the SPP.

MATERIALS

The same SPP materials from Study 1 were used. Differences in prime and target lengths, frequencies, and baseline RTs (RTs for the word presented in isolation) can influence priming effects (Hutchison et al., 2008). So we compared the mean lengths, frequencies (logarithmic HAL frequencies or logHAL), and baseline RTs (taken from the ELP) for both the primes and targets across the three relations using a One-Way ANOVA with Tukey HSD *post-hoc* tests. Neither prime lengths, $F_{(2,\ 249)} = 1.10$, $p = 0.33$, nor target lengths, $F_{(2,\ 249)} = 1.40$, $p = 0.25$, differed across the three relations. However, prime frequencies differed, $F_{(2,\ 249)} = 14.98$, $p < 0.001$, with reliably greater frequencies for the integrative primes ($M = 9.21$, SD $= 1.59$) compared to the thematic ($M = 8.20$, SD $= 1.45$), $p < 0.001$, and taxonomic primes ($M = 8.06$, SD $= 1.48$), $p < 0.001$, which did not differ. Target frequencies also differed among the three relations, $F_{(2,\ 249)} = 12.57$, $p < 0.001$. Integrative target frequencies ($M = 9.78$, SD $= 1.65$) were

Table 1 | Study 1, Means, Standard Deviations, Minimums, and Maximums of measures and ELP control variables.

	Integrative				Thematic				Taxonomic			
	M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max
logGoogle	6.59	0.93	3.92	9.43	5.06	0.88	2.76	7.45	5.53	0.87	2.73	8.11
LSA	0.23	0.16	0.00	0.83	0.33	0.23	−0.04	0.83	0.38	0.22	−0.01	0.92

Notes: Prime and target frequencies and baseline RTs taken from the English Lexicon Project.

greater than thematic ($M = 9.14$, SD $= 1.28$), $p < 0.01$, which in turn were marginally greater than the taxonomic targets ($M = 8.68$, SD $= 1.38$), $p = 0.10$. Baseline prime RTs differed among the three relations, $F_{(2, 249)} = 5.67$, $p < 0.01$, with faster RTs for the integrative primes ($M = 627$, SD $= 55$) than the thematic ($M = 652$, SD $= 68$), $p < 0.001$, and taxonomic primes ($M = 657$, SD $= 70$, $p < 0.001$), which did not differ, $p = 0.83$. Baseline RTs for the integrative targets ($M = 612$, SD $= 50$) did not differ from the thematic targets ($M = 620$, SD $= 56$), $p = 0.57$, but were marginally faster than the taxonomic targets ($M = 630$, SD $= 61$), $p = 0.08$. Baseline RTs did not differ between the thematic and taxonomic targets, $p = 0.57$. Given these differences, we next assessed whether prime frequencies, target frequencies, baseline prime RTs, and baseline target RTs were associated with our primed target RT at each SOA. Correlations with the primed target RTs at each SOA were reliable for only the target frequencies ($r = 0.42$ and $r = 0.33$ for the 200 and 1200 ms SOAs, $ps < 0.001$) and baseline target RTs ($r = -0.35$ and $r = -0.25$ for the 200 and 1200 ms SOAs, $ps < 0.001$), so we included these two variables as covariates in our analyses below. As discussed in the Introduction, we did not predict any differences among word recognition latencies or priming effects for our weakly associated items at either SOA.

RESULTS AND DISCUSSION

We conducted two separate 3 (Relation: integrative, thematic, taxonomic; between-items) × 2 (SOA: 200, 1200; within-items) mixed ANCOVAs on the target RTs and PEs with target frequencies and baseline (ELP) target RTs as covariates. Adjusted mean RTs and PEs for each relation are shown in **Table 2**. Contrary to the results of Sachs et al. (2008), we found equivalent target RTs, $F_{(2, 245)} = 1.82$, $p = 0.17$, and priming effects, $F < 1$, $p = 0.82$, across the three relations. The lack of difference among the relations was consistent across both SOAs, as evident by the lack of an interaction for the RTs, $F < 1$, $p = 0.92$, and PEs, $F_{(2, 247)} = 1.28$, $p = 0.28$, nor was there an effect of SOA for either RTs, $F < 1$, $p = 0.78$, or PEs, $F_{(2, 247)} = 1.03$, $p = 0.31$. Not surprisingly, the target frequencies and baseline target RTs had a reliable effect on RTs ($ps < 0.001$), but did not impact PEs ($ps > 0.45$). No other covariates or interactions were reliable.

One-sample t-tests revealed reliable PEs (>0) for all relations at the 200 ms SOA ($ps = 0.01$). However, at the 1200 ms SOA, only the taxonomic items had reliable priming effects ($p = 0.01$), whereas the thematic and integrative items did not ($p = 0.15$ and $p = 0.54$, respectively). The effects for the taxonomic items are consistent with prior studies (e.g., McRae and Boisvert, 1998; Estes and Jones, 2009; for reviews see Neely, 1991; Jones and Estes, 2012) showing the rapid emergence of taxonomic priming and either the maintenance or an increase of priming magnitudes with increasing SOAs up to 1500 ms. Unfortunately, far fewer studies have investigated the maintenance of PEs for integrative and thematic items in a standard LDT with long SOAs. Estes and Jones (2009) found reliable PEs for integrative items at long SOAs of 1500, 2000, and 2500, and Jones et al. (2011) found larger PEs for integrative, thematic, and taxonomic relations at a 2000 ms SOA than at a short 100 ms SOA. However, in both of those studies, priming effects were based on the difference in target RTs following related versus non-linguistic and repetitive neutral primes (********). Such neutral primes tend to artificially inflate the RTs for the control condition at long SOAs, which in turn yield inflated priming effects (e.g., De Groot et al., 1982; Jonides and Mack, 1984; Jones, 2012).

These results fail to replicate the finding by Sachs et al. (2008) of different priming effects for thematic versus taxonomic items at a 200 ms SOA. Although there were no reliable differences in RTs or PEs among the relations at the 1200 ms SOA, only the taxonomic items had a reliable priming effect. The lack of priming at this longer 1200 ms SOA for the integrative and thematic items seems to preclude expectancy generation as an underlying mechanism. Indeed, the results of Jones (in preparation) suggest that strong FAS is required for integrative priming to occur for longer SOAs >1500 ms. Likewise, thematic priming for strongly associated versus weakly associated pairs may show a similar pattern with reliable priming for only the strongly associated pairs at long SOAs >1500 ms. In contrast, taxonomic priming is often attributed to semantic matching (Neely, 1991) or post-lexical integration (De Groot, 1984, 1985) which entails a search for a plausible relation between prime and target. Categorical relations would be particularly strong for our subject population of young adults attending a university (for review see Estes et al., 2011), and consequently may be better maintained in working memory over long SOAs than the integrative and thematic relations.

Finally, the inclusion of different targets across the three relations in this study and in Sachs et al. (2008) is less than ideal despite the equating or controlling of the confounding variables of target frequencies and baseline target RTs. Hence, in Studies 3 and 4, we develop a set of items so that each target (e.g., *book*) is paired with an integrative (e.g., *instruction*), thematic (e.g., *editor*), and taxonomic (e.g., *article*) prime.

Table 2 | Study 1, Adjusted Means and (SEs) of Target RTs (ms) and Priming Effects (ms).

Relation	200 ms SOA		1200 ms SOA	
	RT	Priming effect	RT	Priming effect
Integrative	670 (6)	27***	699 (7)	6
Thematic	660 (7)	23**	684 (7)	12
Taxonomic	664 (6)	18*	687 (7)	24**

*Notes: Priming Effect = Unrelated RT − Related RT. *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001.*

OVERVIEW OF STUDIES 3 AND 4

The primary purpose of Studies 3 and 4 was to replicate and extend the results found in Studies 1 and 2 using a more controlled set of items having the same target across each relation. We begin with an item analysis to further demonstrate distinct patterns on the co-occurrence measures of LSA and logGoogle across the three relations (Study 3). As mentioned in the Introduction, we extend this item analysis to also include BEAGLE cosines, feature similarity ratings, familiarity ratings. We also include our relation defining measures of relational integration, thematic relatedness relations, and category co-membership in order to verify our classification into relational categories. Next we investigate the extent to which these measures differentially predict lexical priming across the three relations using a standard LDT with 100, 500, and 800 ms SOAs (Study 4).

STUDY 3

As in Study 1, we minimized and equated association strength and assessed local co-occurrence or word pair frequency (Google hits), and global co-occurrence (LSA cosines and BEAGLE cosines). In addition to these database and computational measures, a total of 130 Wayne State University undergraduates provided ratings for categorical relatedness, thematic relatedness, integration, feature similarity, and familiarity. Each of these additional measures is described in detail below along with the relevance to semantic priming theories and the predicted differences across the three relations.

BEAGLE COSINES

The Bound Encoding of the Aggregate Language Environment (BEAGLE; Jones and Mewhort, 2007), also predicts lexical priming. Like the compound-cue model, it attributes lexical priming to the co-occurrence between prime and target. BEAGLE cosines represent a measure of the degree of shared contexts between prime and target. Pairs that are both associative and semantic (i.e., co-occurring and similar in meaning; e.g., *nurse—doctor*) are predicted to have higher BEAGLE cosines than those that are only associative (e.g., *bee—honey*) or only semantic (e.g., *deer—pony*). BEAGLE incorporates both "co-occurrence information" (i.e., information about the word's context) and "transition information" (i.e., information about a word relative to other words in a context such as the intervening words; (Jones and Mewhort, 2007, p. 5). So whereas LSA captures both similarity and textual or global co-occurrence, BEAGLE goes a step further by additionally representing transition information. Thus, given the multi-dimensional aspect of BEAGLE, cosines may be consistent across our three relations.

FEATURE SIMILARITY

The features that we attend to in objects and concepts are likely to be those that help us do things like select appropriate actions and solve problems. The relation (integrative, taxonomic, thematic) between two concepts is partially determined by the distribution of common features among the items (i.e., feature similarity). Taxonomic categories are based on common features among category members (e.g., Rosch, 1975; Markman and Wisniewski, 1997). It is inherent that taxonomic category members have common properties (high feature similarity)—if they did not then taxonomic category membership could not guide particular types of inference and action in the face of incomplete information. Feature similarity can also influence the occurrence and extent of lexical priming, particularly at shorter SOAs. For instance, McRae and Boisvert (1998) found that reliable lexical priming occurred for their highly similar pairs (e.g., *goose—turkey*) at a 250 ms SOA but not for the less similar pairs (e.g., *robin—turkey*). Thematically related items are often based on the ability of the items to play complementary roles in the same scenario (Lin and Murphy, 2001), which is facilitated (but not necessitated) by items having non-overlapping features (e.g., *cake—ice-cream* is more thematically related than *cake—pie*). However, many thematically related pairs (e.g., *prescription—doctor*) are based primarily on their complementary roles in the same event and are not dependent on the extent of overlapping features between items. Item pairs that share very few common features are possible candidates for integrative relations. Integrating two concepts into a single, modified concept requires very low overlap in features between items (Estes, 2003a).

As in Estes and Jones (2009), participants ($N = 25$) rated the feature similarity of each word pair on a scale from 1 (not at all similar) to 7 (very similar). Feature similarity was emphasized in the instructions and differentiated via examples from association and co-occurrence. Instructions for this and all subsequent rating tasks are included in Appendix B. Based on the prior research described above, we predicted that feature similarity would be highest for the taxonomic pairs and lowest for the integrative pairs with the thematic pairs having a feature similarity intermediate between these two other relations.

FAMILIARITY RATINGS

As an additional measure of local co-occurrence or word pair frequency, participants ($N = 21$) rated the familiarity for each pair on a scale from 1 (unfamiliar) to 7 (very familiar). We also assessed the familiarity of our prime-target pairs. As previously mentioned, familiarity is moderately correlated with Google hits. In addition to highly frequent word pairs, familiarity is also likely to be high for words that seem to go together in a given event (e.g., *party—cake*). Hence, we predict higher familiarity ratings for the integrative and thematic items than for the taxonomic pairs.

RELATION VERIFICATION RATINGS

In order to select a final set of the most representative items possible for each relation and to verify our designation of each word pair as taxonomic, thematic, or integrative, we collected category co-membership, thematic relatedness, and integrative ratings, respectively. In making our selection of items to include in the final set, we adopted the criteria that the rating measure should be equal to or greater than the midpoint of 4.00 (on a scale of 1–7) for the respective measure representing that relation (e.g., all thematic items should have a thematic relatedness rating of 4 or greater). Additionally, each of the three measures should be reliably higher for the items in the represented relation than for the items in the other two relations (e.g., thematic relatedness ratings should be reliably higher for the thematic than for the taxonomic or integrative items). For each of the following three rating tasks, the 60 targets were presented with each of their prime-types and the presentation order of all 180 items was randomized across participants.

Categorical co-membership ratings

Because category membership is based on more than just feature similarity (e.g., Spalding and Ross, 2000), we needed to directly assess the extent to which each prime and target belonged to the same specific taxonomic category. Participants ($N = 28$) rated each pair from 1 (not at all category co-members) to 7 (definitely co-members of the same specific category). Instructions distinguished taxonomic relatedness over thematic relatedness and relational integration by emphasizing co-membership in a specific category (see Appendix B).

Thematic relatedness ratings

Participants ($N = 27$) rated the extent to which each pair of concepts was linked together in a common scenario, event, or function on a scale from 1 (not thematically connected) to 7 (highly thematically connected). Instructions emphasized that thematically related concepts were often not featurally similar (see Appendix B).

Relational integration ratings

To better distinguish integrative relations from thematic relations we used the sentential integrative rating task from Estes and Jones (2009), which was found in that study to be highly correlated with integrative ratings for the isolated word pair ($r = 0.80$). Participants ($N = 29$) rated the extent to which the word pair made sense as an object within a sentential context from 1 (not at all sensible) to 7 (completely sensible). The same sentence frame was used for each target across the three relations with the word pair shown in ALL CAPS as the object of each sentence (e.g., "Irene ordered the CHILI SOUP"—taxonomic; "Irene ordered the BOWL SOUP"—thematic; "Irene ordered the TOMATO SOUP"—integrative). Note that in this integrative rating task, the integrative pairs (e.g., *tomato soup*) should have much higher ratings than the thematic pairs, which are not as readily integrative (e.g., *bowl soup* does not easily denote a subtype of soup, as soup is typically served in a bowl).

MATERIALS

Based on the results from the three relational verification rating tasks, we narrowed down the prior set of 180 items (60 per relation) to a final set of 132 items (44 per relation) in order to better minimize the degree to which items could represent more than one relation. This final set of items is shown in Appendix C.

RESULTS AND DISCUSSION

The means, SDs, minimums, and maximums on each of these measures (5 rating tasks and 3 computational measures) are shown for each relation in **Table 3**. Separate One-Way ANOVAs and LSD *post-hoc* tests (see **Table 4**) with the relation representative measures of integrative ratings, thematic relatedness, and category co-membership confirm that: (1) the integrative items had higher integrative ratings than did the taxonomic and thematic items, (2) the thematic items had higher thematic relatedness ratings than the integrative and taxonomic items, and (3) the taxonomic items had higher category co-membership ratings than the other two relations. Moreover, as shown in **Table 4**, separate One-Way ANOVAs on the remaining measures revealed reliable differences among the three relations for feature similarity ratings, LSA, and familiarity ratings, but only marginally for logGoogle, and not for BEAGLE. Unsurprisingly, feature similarity

Table 3 | Study 3, Means, Standard Deviations, Minimums, and Maximums of Measures.

Measure	Integrative				Thematic				Taxonomic			
	M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max
Integrative ratings	5.50	0.82	4.00	6.69	4.79	1.14	2.97	6.76	3.81	1.04	1.97	6.28
Thematic relatedness	4.69	0.82	2.93	6.33	5.59	0.63	4.19	6.52	4.92	0.67	3.85	6.30
Category co-members	3.74	0.74	2.25	5.68	4.37	0.63	3.00	5.57	4.95	0.54	4.00	6.00
Feature similarity	3.36	0.74	2.00	5.08	3.78	0.66	2.60	5.48	5.37	0.53	4.16	6.28
Familiarity	5.63	0.79	3.67	6.76	5.97	0.45	4.90	6.81	5.14	0.74	3.24	6.24
logGoogle	5.21	0.67	2.71	6.08	5.08	0.49	3.47	5.88	4.95	0.53	3.73	6.11
BEAGLE	0.27	0.14	0.02	0.66	0.29	0.15	−0.01	0.71	0.31	0.17	0.04	0.65
LSA	0.27	0.18	0.04	0.79	0.38	0.19	0.07	0.72	0.39	0.19	0.09	0.75

Table 4 | Study 3, differences among relations for each measure.

Measure	ANOVA	Comparisons (*Post-hoc*, LSD)
Integrative ratings	$F = 39.02$, $p < 0.001$	Integrative > Thematic > Taxonomic
Thematic relatedness	$F = 31.20$, $p < 0.001$	Thematic > (Taxonomic = Integrative)
Category co-members	$F = 18.94$, $p < 0.001$	Taxonomic > Thematic > Integrative
Feature similarity	$F = 117.12$, $p < 0.001$	Taxonomic > Thematic > Integrative
Familiarity	$F = 16.36$, $p < 0.001$	Thematic > Integrative > Taxonomic
logGoogle	$F = 2.38$, $p < 0.10$	**Integrative > Taxonomic,** Integrative = Thematic, Taxonomic = Thematic
BEAGLE	$F < 1$, $p = 0.42$	Taxonomic = Thematic = Integrative
LSA	$F = 5.75$, $p < 0.01$	**(Taxonomic = Thematic) > Integrative**

*Notes: Comparisons shown in **bold** font replicate results found in Study 1.*

ratings were higher for the taxonomic items than for the thematic items, which in turn were higher than those for the integrative items. As in Study 1, LSA cosines were lower for the integrative items in comparison to the taxonomic and thematic items, which were equivalent. In addition to reflecting similarity of the taxonomic items, the high LSA cosines may have simply reflected the fact that members of a given category often co-occur within the same text. Familiarity ratings were higher for the thematic than the integrative items, which in turn were higher than the taxonomic ratings. Word pair frequencies (logGoogle) hits were higher for the integrative items than the taxonomic items, but were equal to the thematic items. The lack of difference between the thematic and integrative items may reflect the ability to integrate several of the thematic pairs into a sensible entity (e.g., *lab coat, jelly jar*).

Predictor variable inter-correlations

The inter-correlations among these measures for all 132 items are shown in **Table 5**. In the next few sub-sections, we highlight some of the correlations that show further distinction across our three relations.

Inter-correlations with integrative ratings

Despite the overlap between integrative and thematic relations in general and for some of our items (e.g., *ambulance siren, shower soap*), we found no overall relationship between integrative and thematic ratings across our item set. The integrative ratings and category co-member (i.e., taxonomic) ratings were inversely related. Likewise, the inverse relationships between feature similarity and integrative ratings across all items are consistent with the dissociation between integrative (a.k.a., "relational") and taxonomic (a.k.a., "attributive") pairs observed in lexical priming (Estes and Jones, 2009, Experiment 2) and conceptual combination (Wisniewski and Love, 1998; Estes, 2003b) studies. For instance, across the 45 integrative and 45 "semantic" (i.e., taxonomic) items used by Estes and Jones, there was an inverse relationship between the sentential integrative ratings and feature similarity ratings ($r = -0.55$, $p < 0.001$). These inverse correlations further underscore the difficulty (but not impossibility) of relationally integrating two highly similar items from the same category (e.g., *cow horse, lake ocean, knife spoon*). Yet, as mentioned in the Introduction, there is also overlap between taxonomic and integrative relations. Despite our best efforts to tease apart the three relations in the creation of our item set, this overlap was reflected by a few items of our taxonomic and integrative pairs (e.g., *alarm siren*, *pork bacon*, *suit pants, chocolate candy*) that had high ratings across category co-membership, feature similarity, and integration. These items likely reduced the extent to which integrative ratings were inversely correlated with category co-membership and feature similarity. As shown in **Table 5**, integrative ratings were positively and robustly associated with familiarity, though only weakly related to logGoogle hits. However, integrative ratings were inversely related to the more global co-occurrence measures of BEAGLE and LSA cosines.

Inter-correlations with thematic relatedness ratings

In contrast, thematic relatedness was positively associated with category co-membership. This positive association is consistent with Lin and Murphy (2001), who argued that thematic relations (e.g., *chalk/blackboard*) sometimes create more coherent categories than taxonomic relations (e.g., *chalk/marker*). However, as demonstrated by the merely marginal correlation between thematic relatedness and feature similarity, members of thematic categories do not cohere around shared features. Rather, members of thematic categories are united by playing complementary roles in the same scenario or event (Estes et al., 2011). The correlation between thematic relatedness and feature similarity is relatively weak because objects that have the same properties and affordances are unlikely to be able to engage in a complementary action (although for some exceptions see Wisniewski and Bassok, 1999). Thematic ratings were also robustly correlated with familiarity but not with logGoogle or BEAGLE. Hence, subjective familiarity reflects not only the ability to integrate two concepts, but also (and to a slightly greater degree) the co-occurrence of the concepts within an event. However, in contrast to the inverse correlation with the integrative ratings, LSA cosines were positively associated with thematic relatedness. Hence, the respective correlations with LSA cosines further distinguish between thematic and integrative relations.

Inter-correlations with category co-membership ratings

Not surprisingly, category co-membership was strongly and positively associated with feature similarity ratings. This robust

Table 5 | Study 3, Inter-correlations of ratings and computation measures.

	1	2	3	4	5	6	7	8
Integrative ratings	–	–	–	–	–	–	–	–
Thematic relatedness	0.04	–	–	–	–	–	–	–
Category co-members	−0.25**	0.60***	–	–	–	–	–	–
Feature similarity	−0.46***	0.17†	0.74***	–	–	–	–	–
Familiarity	0.50***	0.63***	0.22*	−0.16†	–	–	–	–
logGoogle	0.18*	0.09	−0.04	−0.06	0.33***	–	–	–
BEAGLE	−0.21*	0.08	0.20*	0.18*	0.08	0.42***	–	–
LSA	−0.19*	0.30***	0.36***	0.26**	0.14	0.15†	0.49***	–

Notes: $^{\dagger}p < 0.10$, $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$.

correlation is consistent with models of categorization and prior studies. Family resemblance approaches to category coherence are based on the tenet that taxonomic categories cohere around common features (Rosch, 1975; Rosch and Mervis, 1975). The importance of feature similarity to category structure is reflected in the relationship between category membership and perceived similarity. Category co-members like *milk* and *lemonade* are regularly judged to be more similar to one another than category non-members like *milk* and *horse* (Murphy and Brownell, 1985; Wisniewski and Bassok, 1999; Golonka and Estes, 2009). Thus, category membership was strongly related to similarity. In contrast to the integrative and thematic ratings, category co-membership was only weakly related to familiarity. Consistent with the thematic relatedness ratings, category co-membership was not related to logGoogle hits but was reliably related to LSA cosines. In contrast to the inverse correlation with integrative ratings, and the lack of an association with the thematic ratings, BEAGLE cosines were related (albeit weakly) to category co-membership.

Inter-correlations among the co-occurrence measures, similarity ratings, and familiarity

Though not a primary goal of our study, we briefly highlight some of the inter-correlations that replicate interesting patterns found in prior studies. As discussed in Study 1, it is increasingly common to use LSA cosines as a proxy for similarity. However, like Simmons [Golonka] and Estes (2006), we found only a weak association between LSA cosines and feature similarity ratings. In support of our claim that LSA is a better measure of textual co-occurrence than similarity, LSA cosines were more strongly correlated with BEAGLE ($r = 0.49$) in comparison to feature similarity ratings ($r = 0.26$). This finding also corroborates the results of Maki and Buchanan's (2008) exploratory factor analysis, which found that LSA along with BEAGLE more strongly loaded on the text-based factor rather than the similarity factor. As with LSA, BEAGLE cosines were only weakly related to feature similarity ratings.

In direct contrast to LSA cosines, logGoogle hits were reliably related to integrative ratings but not to thematic relatedness or category co-membership ratings. Also, in direct contrast to the two more global co-occurrence measures (LSA and BEAGLE cosines), logGoogle was not related to feature similarity ratings. The three co-occurrence measures (logGoogle, BEAGLE, and LSA) were interrelated, though to a much lesser extent between logGoogle and LSA. The correlation between BEAGLE and LSA is consistent with that found by Jones and Mewhort (2007, **Table 5**; $r = 0.37$). Moreover, familiarity ratings were related to logGoogle, consistent with the findings of Wisniewski and Murphy (2005), but not to BEAGLE or LSA. The finding that BEAGLE was more related to LSA and to logGoogle (both rs > 0.40) than these two measures were to each other indicate that BEAGLE cosines reflect both local and global co-occurrence. Indeed, this finding supports the BEAGLE model's incorporation of both "co-occurrence information" (i.e., information about the word's context) and "transition information" (i.e., information about a word relative to other words in a context; Jones and Mewhort, 2007, p. 5). In Study 4, we predict that the various co-occurrence measures (logGoogle, LSA cosines, BEAGLE) will differentially predict lexical priming across the three relations.

STUDY 4

As shown in Study 3, global measures of co-occurrence (LSA and BEAGLE) were particularly high for both the taxonomic and thematic pairs. For these items, we predict that the more global co-occurrence measures should facilitate priming effects by facilitating global integration (Chwilla and Kolk, 2005), or expectancy processing, in which an upcoming target is anticipated based on its frequent inclusion in an event (McRae and Matsuki, 2009; Metusalem et al., 2012). For instance, Chwilla and Kolk attributed lexical priming in a LDT with a short SOA for target items following two simultaneously script-related primes (e.g., *move—piano* → *backache*) to their global integration model and to higher LSA cosines for their script-related items than their unrelated items. A similar study (Khalkhali et al., 2012) attributed lexical priming for targets following individually presented primes depicting events that occurred prior to the target event (e.g., *marinate* → *grill* → *chew*) to the integration of the prime concepts into a situation model (i.e., a mental representation of a sequence of events). As with Chwilla and Kolk (2005), LSA cosines were also higher for the related than the unrelated triplets. Likewise, Jones and Mewhort (2007) found that BEAGLE cosines predicted priming for the semantic (mostly taxonomic) non-associative pairs (e.g., *deer—pony*) used in Chiarello et al. (1990). So then, these findings tentatively suggest that global co-occurrence (LSA and BEAGLE cosines) may predict target word recognition latencies following thematic and taxonomic primes.

For the integrative items, word pair frequencies (logGoogle hits) should predict lexical priming, particularly at short SOAs. The Embodied Conceptual Combination (ECCo) model (Lynott and Connell, 2010) posits a "quick and dirty" linguistic shortcut in which interpretation times (and by extension word recognition times) are faster for more frequently co-occurring combinations. This theory of conceptual combination interpretation is congruent with the compound-cue theory in lexical priming (Ratcliff and McKoon, 1988; McKoon and Ratcliff, 1992) which argues that prime-target compounds that are highly co-occurring in long-term memory produce faster RTs than less accessible ones. Hence, based on the ECCo and compound cue theories, logGoogle should influence target RTs, but only at the short 100 ms SOA.

METHOD

Participants

Wayne State University undergraduates ($N = 223$) participated for partial course credit and were randomly assigned to the 100 ms SOA ($n = 57$), the 500 ms SOA ($n = 105$) or the 800 ms SOA ($n = 61$).

Materials

Experimental items consisted of the final set of items from Study 3 (see Appendix C). As in Study 2, prime frequencies (logHAL), length, and RTs were taken from the ELP website (Balota et al., 2007, http://elexicon.wustl.edu/) and compared across the three relations. A One-Way ANOVA found no reliable differences among the relations for prime length, $F < 1$, $p = 0.78$, or prime RTs, $F_{(2, 126)} = 1.05$, $p = 0.35$. However, prime frequencies differed among the relations, $F_{(2, 126)} = 3.12$, $p < 0.05$, with reliably greater frequencies for the integrative primes ($M = 9.10$, SD = 1.49) and marginally greater frequencies for the thematic primes ($M = 8.92$, SD = 1.30) in comparison to the taxonomic primes ($M = 8.33$, SD = 1.68).

Procedure

Participants responded only to the target words. On each of four experimental lists, critical trials consisted of 44 real word targets following an integrative prime (11 trials), thematic prime (11 trials), taxonomic prime (11 trials), or an unrelated prime (11 trials). An additional 44 filler trials consisted of a real word prime followed by a non-word target (e.g., *page—hife*). As in prior studies (e.g., Jones, 2012), non-word primes were selected from the ELP (Balota et al., 2007) so that they would not differ in length from the real word primes. Prime-types were counterbalanced across lists. Primes were vertically and horizontally centered in 22-point red Arial font on a black screen and targets were in white font. Participants pressed the spacebar to begin each trial. A blank screen appeared for 200 ms followed by a fixation plus sign (+) for 500 ms. Next the prime word appeared for 100 ms immediately followed by the target in the 100 ms SOA condition or by a blank screen for 400 ms in the 500 ms SOA or 700 ms in the 800 ms SOA. Targets remained on the screen until participants indicated whether the letter string was a real word by pressing the J key for "yes" or the F key for "no." A 1000 ms inter-trial interval separated each trial, and presentation order of the 88 trials was randomized across participants. Ten practice trials preceded the 88 experimental trials.

RESULTS AND DISCUSSION

RTs from incorrect trials (1.4% of the data) were excluded from analyses in addition to RTs greater than 1500 ms and any remaining RTs greater than 2.5 SDs above or below each participant's condition mean (an additional 5.6%). Mean response times and accuracies were analyzed using a 3 (SOA: 100, 500, 800; between-participants) ×4 (Prime-type: integrative, taxonomic, thematic, unrelated; within-participants) ANOVA across participants F_p and items F_i. All factors were within items. Accuracies were at ceiling (all Ms = 0.98) and there were no reliable main effects or interactions ($p > 0.20$).

Mean RTs and priming effects are shown in **Table 6**. Overall, RTs were slower for targets following the unrelated primes than for targets following the three related primes, $F_{p(3, 660)} = 11.96$, $p < 0.001$, $\eta_p^2 = 0.05$, and $F_{i(3, 129)} = 10.28$, $p < 0.001$, $\eta_p^2 = 0.19$. There was no main effect of SOA by subjects, $F_{p(2, 220)} = 1.68$, $p = 0.19$. Within the item analysis, RTs did not differ between the 500 ms and 800 ms SOAs (both Ms = 667), but were faster than in the 100 ms SOA ($M = 701$), $F_{i(2, 86)} = 50.69$, $p < 0.001$. There was not an interaction between SOA and Relation, $F_{p(6, 660)} < 1$, $p = 0.51$, and $F_{i(6, 258)} = 1.35$, $p = 0.24$.

To determine whether there were differences in priming magnitude among the integrative, thematic, and taxonomic relations, we ran a 3 (SOA) × 3 (Relation) ANOVA with priming effects (unrelated RTs—related RTs) as the dependent measure. Priming effects did not differ among the three relations, $F_{p(2, 440)} < 1$, $p = 0.56$, and $F_{i(2, 86)} < 1$, $p = 0.54$, or among the SOAs, $F_{p(2, 220)} < 1$, $p = 0.86$, and $F_{i(2, 86)} < 1$, $p = 0.88$, nor was there a reliable interaction, $F_{p(4, 440)} = 1.28$, $p = 0.28$, and $F_{i(4, 172)} = 1.96$, $p = 0.10$. The consistent priming effects among all three relations across the SOAs replicates the pattern of results found in Study 2 and in our earlier study using neutral primes (Jones et al., 2011). Hence, in contrast to the results of Sachs and colleagues (2008), thematic priming was not more robust than taxonomic priming.

Partial correlations and regression analyses

Though priming effects were equivalent among the three relations within each SOA, different underlying measures were related to priming within each relation at each SOA (see **Table 7**). Because prime frequencies and prime latencies can influence target RTs (Hutchison et al., 2008), especially at short SOAs, we controlled for prime logHAL frequencies (which differed among the three relations—see "Materials" section) and baseline prime RTs (taken from the ELP). Given the numerous factors that influence target word recognition times (e.g., frequencies, orthographic neighborhoods, etc.), we also included the baseline target RTs (also taken from the ELP) as a control variable. For the partial correlation analyses reported below, we examined the influence of several common factors related to word recognition latencies, namely, feature similarity (e.g., McRae and Boisvert, 1998), LSA (e.g., Hare et al., 2009; Jones, 2012), BEAGLE (e.g., Jones and Mewhort, 2007; Hare et al., 2009), and Google hits (e.g., Jones, 2010, 2012). All marginal ($p < 0.10$) and reliable ($p < 0.05$) predictors found in the partial correlation analyses were then further examined in hierarchical stepwise regression analyses for the applicable SOA's target RTs with the same control variables (prime frequencies, baseline prime RTs, and baseline target RTs) entered into the first

Table 6 | Study 4, Means and (SEs) of RTs (ms), Priming Effects (ms), and Predictors of Target RTs and Priming Effects.

	100 ms SOA		500 ms SOA		800 ms SOA	
Relation	**RT**	**Priming effect**	**RT**	**Priming effect**	**RT**	**Priming effect**
Unrelated	722 (15)	–	689 (15)	–	687 (14)	–
Integrative	700 (14)	22*	671 (13)	18*	654 (13)	33*
Thematic	690 (14)	32**	671 (13)	18*	661 (14)	26**
Taxonomic	694 (16)	26**	655 (13)	34**	660 (15)	27**

*Notes: Priming Effect = Unrelated RT − Related RT. *p ≤ 0.05; **p ≤ 0.01.*

Table 7 | Study 4, partial correlations—predictors of target response times (ms) by relation.

	100 ms SOA	500 ms SOA	800 ms SOA
INTEGRATIVE ITEMS			
Feature similarity	−0.28†	−0.06	−0.18
logGoogle	−0.47**	−0.18	−0.25
BEAGLE	−0.23	−0.15	−0.18
LSA	−0.18	−0.001	−0.15
THEMATIC ITEMS			
Feature similarity	0.06	0.16	−0.31*
logGoogle	−0.02	−0.34*	−0.03
BEAGLE	0.16	−0.16	−0.33*
LSA	0.19	0.04	−0.26†
TAXONOMIC ITEMS			
Feature similarity	−0.17	−0.24	−0.16
logGoogle	−0.11	−0.36*	−0.11
BEAGLE	−0.08	−0.22	−0.32*
LSA	0.19	−0.12	0.02

Controlled for the following variables taken from the English Lexicon Project (ELP): Prime (HAL) logFrequency, Prime RTs, Target RTs. Note: $^{\dagger}p \leq 0.10$; $^{*}p \leq 0.05$; $^{**}p \leq 0.01$.

block and the marginal and reliable correlates entered into the second block. Variables that had beta coefficients with significance levels greater than 0.05 were excluded from the best fitting model.

Integrative items

Local co-occurrence (word pair frequencies) as measured by logGoogle was reliably related to the target RTs at the 100 ms SOA. Additionally, we found a marginal correlation between feature similarity ratings and target RTs at this short SOA. No other variables approached conventional levels of significance. These correlates (logGoogle and feature similarity ratings) and the criterion measure of target RTs for the 100 ms SOA were entered into a hierarchical stepwise regression model with only the control variables entered in the first block and the predictors added to the second block. Results demonstrated that the inclusion of logGoogle ($\beta = -0.42$, $t = 2.71$, $p = 0.01$) along with baseline target RTs ($\beta = 0.34$, $t = 2.18$, $p < 0.05$) as the best fitting model, $R = 0.63$, $R^2 = 0.40$, $F_{(2, 40)} = 13.28$, $p < 0.001$. Moreover, the addition of logGoogle was reliably more predictive of target RTs in the 100 ms SOA than only the baseline target RTs in Model 1, $\Delta R^2 = 0.17$, $F_{(1, 40)} = 11.04$, $p < 0.01$. Either compound-cue theory or the ECCo theory may explain these effects at this short 100 ms SOA. That is, the more familiar or frequently co-occurring the word pair, the easier it is to retrieve the representation of that entity from memory. The current results suggest a very rapid process of retrieval consistent with the ECCo model and compound-cue theories.

Thematic items

In contrast to the results for the integrative items, no predictors were reliably related to target RTs at the 100 ms SOA. Word pair frequencies (logGoogle) were related to target RTs at only the 500 ms SOA but did not approach significance at the shorter 100 ms or the longer 800 ms SOAs. Interestingly, the co-occurrence measures (LSA and BEAGLE) and feature similarity were related to word recognition target RTs at the 800 ms SOAs (albeit only marginally for LSA) but not at the 100 or 500 ms SOAs. As before, within the applicable SOAs, we included the marginal and reliable correlates along with the control variables in the hierarchical stepwise regression analyses to determine whether these co-occurrence variables explained the variance in target RTs above and beyond the control variables. Within the 500 ms SOA, the best fitting model included baseline target RTs ($\beta = 0.35$, $t = 2.63$, $p = 0.01$), baseline prime RTs ($\beta = 0.32$, $t = 2.39$, $p < 0.05$), and logGoogle ($\beta = -0.28$, $t = 2.16$, $p < 0.05$), $R = 0.58$, $R^2 = 0.33$, $F_{(3, 39)} = 6.50$, $p = 0.001$. The addition of logGoogle to the model explained more of the variance than just the two control variables alone, $\Delta R^2 = 0.08$, $F_{(1, 39)} = 4.69$, $p < 0.05$. Within the 800 ms SOA, the best fitting model included only the baseline target RTs ($\beta = 0.36$, $t = 2.59$, $p = 0.01$) and BEAGLE cosines ($\beta = -0.31$, $t = 2.26$, $p < 0.05$), $R = 0.54$, $R^2 = 0.29$, $F_{(2, 40)} = 8.30$, $p = 0.001$. Moreover, the inclusion of BEAGLE cosines accounted for more of the variance than baseline target RTs alone, $\Delta R^2 = 0.09$, $F_{(1, 40)} = 5.11$, $p < 0.05$.

Our results for the 800 ms SOA corroborate those of prior studies (Chwilla and Kolk, 2005; Hare et al., 2009), which also found an influence of global co-occurrence (LSA and BEAGLE) for most thematic relations. The influence of global co-occurrence measures like BEAGLE on target RTs reflects the activation of event knowledge, because words that are related to a common event co-occur (Hare et al., 2009). Notably, only word pair frequencies (logGoogle) were predictive beyond the control variable at the 500 ms SOA, which suggests an initial attempt at a more local integration between the two concepts prior to a more global integration of the two concepts within an event at the 800 ms SOA. The time course of activation of such event knowledge in our standard LDT is consistent with that found in other word recognition studies (e.g., Chwilla and Kolk, 2005). For instance, Chwilla and Kolk found faster RTs and an N400 priming effect for targets following two non-associated script related primes (e.g., *backache* following the simultaneously presented primes *move* and *piano*). These primes were presented for 400 ms and the N400 effect occurred an additional 400–500 ms following target presentation for a total duration of approximately 800 ms following prime onset. Hence, our results are consistent with the global integration model proposed by Chwilla and Kolk (2005) or formation of a situation model (Khalkhali et al., 2012), in which global co-occurrence rather than local co-occurrence facilitates the integration of prime and target. Alternatively, expectancy generation (Metusalem et al., 2012) may also account for our results via the formation (given ample time) of a small set of anticipated event-related targets prior to target presentation (e.g., *bacon, breakfast, toast* following *eggs*). Most importantly, these results are the first to demonstrate a key difference in the underlying influences of lexical priming between integrative pairs (e.g., *turkey bacon*) and thematic pairs (e.g., *eggs bacon*).

Taxonomic items

The pattern of results for the taxonomic items across the three SOAs was somewhat similar to that of the thematic items. Within the 500 ms SOA, only word pair frequencies (logGoogle) were reliably correlated with target RTs. Yet in contrast to the thematic items, only the BEAGLE cosines were reliably related to target RTs within the 800 ms SOA. In the regression analyses, the pattern of results was identical to that for the thematic items. Within the 500 ms SOA, the best fitting model included baseline target RTs ($\beta = 0.55$, $t = 4.57$, $p < 0.001$) and logGoogle, ($\beta = -0.29$, $t = 2.44$, $p < 0.05$), $R = 0.67$, $R^2 = 0.45$, $F_{(2, 40)} = 16.09$, $p < 0.001$. The addition of logGoogle accounted for additional variance in target RTs, $\Delta R^2 = 0.08$, $F_{(1, 40)} = 5.96$, $p < 0.05$. Within the 800 ms SOA, the best fitting model included baseline target RTs and BEAGLE cosines, $R = 0.43$, $R^2 = 0.18$, $F_{(2, 40)} = 4.47$, $p < 0.05$. The addition of BEAGLE cosines accounted for additional variance in target RTs, $\Delta R^2 = 0.08$, $F_{(1, 40)} = 4.01$, $p = 0.05$. Moreover, in this model the BEAGLE cosines ($\beta = -0.30$, $t = 2.00$, $p = 0.05$) were a reliable predictor, whereas the baseline target RTs were not ($\beta = 0.22$, $t = 1.46$, $p = 0.15$).

Results corroborate Jones and Mewhort's (2007) finding that BEAGLE cosines predicted priming for non-associated taxonomic items. The correlation between feature similarity and taxonomic target RTs was not reliable in any of the SOAs, though the trend was in the predicted direction. The lack of a reliable correlation may simply reflect the range restriction for feature similarity across these uniformly and highly similar taxonomic items (see minimums and maximums in **Table 3**).

GENERAL DISCUSSION

We demonstrated distinct patterns of underlying factors (e.g., local and global co-occurrence) among weakly associated integrative, thematic, and taxonomic pairs for a large item set with different targets taken from the SPP (Study 1) and for a smaller more controlled item set having the same target across the relations (Study 3). Most notably, our results were the first to demonstrate a distinction between integrative pairs (e.g., *turkey bacon*) and thematic pairs (e.g., *eggs bacon*), with relatively less distinction between the thematic and taxonomic items. Integrative pairs were lower in global co-occurrence (LSA cosines; cf. Studies 1 and 3) and feature similarity (Study 3) in comparison to thematic and taxonomic pairs. We also found distinct patterns of correlations between each of the relational classification ratings (integrative, thematic, and category co-membership) and the other measures, thereby further distinguishing among these relations (see **Table 5**). The integrative ratings were *inversely* related to category co-membership, feature similarity, and LSA cosines, whereas thematic relatedness ratings were *directly* related to these measures (though only marginally related to feature similarity). This distinction between integrative and thematic relations is an important finding for semantic relation researchers using both behavioral and neuroscience methods. For instance, would additional areas of brain activation result for pairs that are both integrative and thematic? On a related note, could the earlier activation obtained for thematic in comparison to taxonomic pairs in earlier studies (e.g., Sachs et al., 2008; Sass et al., 2009) be partially attributed to the ability to integrate some of the thematic items? For example, weakly associated thematic items that can also be easily integrated, as is the case with many locative relations (e.g., *hospital doctor*), may exhibit distinct priming characteristics, time courses of activation, and/or underlying neural regions of activation in comparison to thematic relations that are less easily integrated (e.g., *prescription doctor*).

Despite the distinct pattern on these measures for each relation we found no differences in overall priming magnitudes among the relations in Studies 2 and 4. Recall that prior studies had previously shown such a dissociation between thematic and taxonomic relations in regards to time course and/or strength of activation (generally with earlier and/or more robust activation for thematic than taxonomic items Sachs et al., 2008; Sass et al., 2009; Mirman et al., 2011; Mirman and Graziano, in press). Though priming effects were equivalent among the three relations, the underlying measures of logGoogle and BEAGLE cosines differentially predicted the observed priming within each relation across the three SOAs in Study 4. In turn, the distinct patterns of predictors across the three relations suggest that different priming mechanisms were responsible for each relation. For the integrative items, local co-occurrence (logGoogle) predicted integrative priming at the 100 ms SOA. This finding suggests a "short-cut" in which the integrated prime-target pair may be retrieved from memory similar to that found in conceptual combination interpretation (Lynott and Connell, 2010) or to the compound-cue model in semantic priming (McKoon and Ratcliff, 1992). Consistent with prior studies (e.g., Chwilla and Kolk, 2005; Khalkhali et al., 2012) word recognition of targets following the thematic primes was due to a global integration process as suggested by the correlations with BEAGLE cosines (and to a marginal extent with LSA) at the longer 500 and 800 ms SOAs. Finally, as in Jones and Mewhort (2007), we also found that BEAGLE cosines predicted taxonomic priming. Given that the BEAGLE model's distributed representation extends beyond shared features to also include abstracted representations such as co-exemplars and category labels, it is perhaps not too surprising that BEAGLE cosines should predict taxonomic priming. In turn, this underlying predictor suggests that taxonomic relations are retrospectively activated following target presentation as posited by the semantic matching model (Neely, 1991) and post-lexical integration model (e.g., De Groot, 1984, 1985). Feature similarity may also facilitate this process (McRae and Boisvert, 1998; Thompson-Schill et al., 1998). Though not significant, there was a trend in the predicted inverse direction between feature similarity and word recognition for the taxonomic targets across all three SOAs. The lack of significance is most likely due to the range restriction for feature similarity among the taxonomic items (i.e., >4.00 on a seven-point scale). With greater variation in similarity among taxonomically related items such as the items used by McRae and Boisvert (1998), feature similarity would likely also be a reliable predictor of target word recognition.

IMPLICATIONS FOR BEHAVIORAL AND CORPUS STUDIES

Association strength poses an additional variable that should be examined further. In our current research, we focused on only weakly associated integrative, thematic, and taxonomic pairs, because the associative boost found in prior studies with taxonomic pairs (Moss et al., 1995; for review see Lucas, 2000; Hutchison, 2003; Jones and Estes, 2012) and integrative pairs (Jones, in preparation) may mask the more subtle differences on other measures among these relations. However, one worthy avenue of pursuit would be to compare and contrast the relative impact association strength has on priming effects within each relation across a variety of LDT paradigms favoring perceptual simulation, spreading activation, or expectancy generation. Such an investigation would require a large set of items having equivalent means and variability of association strength across the three relations. The selection of items from large scale studies provide the advantage of larger item sets with a greater variability in the measures of interest, whereas smaller created item sets having the same targets have the advantage of more experimental control.

REFERENCES

Badham, S. P., Estes, Z., and Maylor, E. A. (2012). Integrative and semantic relations equally alleviate age-related associative memory deficits. *Psychol. Aging* 27, 141–152.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English lexicon project: a user's guide. *Behav. Res. Methods* 39, 445–459.

Balota, D. A., Yap, M. J., Hutchison, K. A., and Cortese, M. J. (2012). "Megastudies: what do millions (or so) of trials tell us about lexical processing?" in, *Visual Word Recognition,* Vol. 2, ed J. S. Adelman (Hove, UK: Psychology Press), 90–115.

Chiarello, C., Burgess, C., Richards, L., and Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: some words do, some words don't …sometimes, some places. *Brain Lang.* 38, 75–104.

Chwilla, D. J., and Kolk, H. H. J. (2005). Accessing world knowledge: evidence from N400 and reaction time priming. *Cogn. Brain Res.* 25, 589–606.

Costello, F. J., and Keane, M. T. (2000). Efficient creativity: constraint-guided conceptual combination. *Cogn. Sci.* 24, 299–349.

De Groot, A. M. B. (1984). Primed lexical decision: combined effects of the proportion of related prime-target pairs and the stimulus-onset asynchrony of prime and target. *Q. J. Exp. Psychol.* 36A, 253–280.

De Groot, A. M. B. (1985). Word-context effects in word naming and lexical decision. *Q. J. Exp. Psychol.* 37A, 281–297.

De Groot, A. M. B., Thomassen, A. J. W. M., and Hudson, P. T. W. (1982). Associative facilitation of word recognition as measured from a neutral prime. *Mem. Cogn.* 10, 358–370.

Downing, P. (1977). On the creation and use of English compound nouns. *Language* 53, 810–842.

Estes, Z. (2003a). A tale of two similarities: comparison and integration in conceptual combination. *Cogn. Sci.* 27, 911–921.

Estes, Z. (2003b). Attributive and relational processes in nominal combination. *J. Mem. Lang.* 48, 304–319.

Estes, Z., Golonka, S., and Jones, L. L. (2011). "Thematic thinking: the apprehension and consequences of thematic relations" in *Psychology of Learning and Motivation*, Vol. 54, ed B. Ross (Burlington: Academic Press), 249–294.

Estes, Z., and Jones, L. L. (2006). Priming via relational similarity: a copper horse is faster when seen through a glass eye. *J. Mem. Lang.* 55, 89–101.

Estes, Z., and Jones, L. L. (2009). Integrative priming occurs rapidly and uncontrollably during lexical processing. *J. Exp. Psychol. Gen.* 138, 112–130.

Gagné, C. L. (2002). Lexical and relational influences on processing of novel compounds. *Brain Lang.* 81, 723–735.

Gagné, C. L., and Shoben, E. J. (2002). Priming relations in ambiguous noun-noun combinations. *Mem. Cogn.* 30, 637–646.

Gagné, C. L., and Spalding, T. L. (2004). Effect of relation availability on the interpretation and access of familiar noun-noun compounds. *Brain Lang.* 90, 478–486.

Gagné, C. L., and Spalding, T. L. (2009). Constituent integration during the processing of compound words: does it involve the use of relational structures? *J. Mem. Lang.* 60, 20–35.

Gagné, C. L., Spalding, T. L., and Ji, H. (2005). Re-examining evidence for the use of independent relational representations during conceptual combination. *J. Mem. Lang.* 53, 445–455.

Golonka, S., and Estes, Z. (2009). Thematic relations affect similarity via commonalities. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1454–1464.

Hare, M., Jones, M., Thomson, C., Kelly, S., and McRae, K. (2009). Activating event knowledge. *Cognition* 111, 151–167.

Howard, M. W., and Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *J. Mem. Lang.* 46, 85–98.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychon. Bull. Rev.* 10, 785–813.

Hutchison, K. A., Balota, D. A., Cortese, M. J., Neely, J. H., Niemeyer, D. P., and Bengson, J. J. (2012). *The Semantic Priming Project: A Web Database of Descriptive and Behavioral Measures for 1, 661 nonwords and 1, 661 English Words Presented in Related and Unrelated Contexts*. (Montana State University). http://spp.montana.edu.

Hutchison, K. A., Balota, D. A., Cortese, M. J., and Watson, J. M. (2008). Predicting semantic priming at the item level. *Q. J. Exp. Psychol.* 61, 1036–1066.

Jones, L. L. (2010). Pure mediated priming: a retrospective semantic matching model. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 135–146.

Jones, L. L. (2012). Prospective and retrospective processing in associative mediated priming. *J. Mem. Lang.* 66, 52–67.

Jones, L. L., and Estes, Z. (2012). "Lexical priming: associative, semantic, and thematic influences on word recognition," in *Visual Word Recognition*, Vol. 2, ed J. S. Adelman (Hove, UK: Psychology Press) 44–72.

Jones, L. L., Estes, Z., and Marsh, R. L. (2008). An asymmetric effect of relational integration on recognition memory. *Q. J. Exp. Psychol.* 61, 1169–1176.

Jones, L. L., Estes, Z., and Golonka, S. (2011). "Dissociation of thematic, taxonomic, and integrative relations," in *Poster Presented at the 33rd Annual Meeting of the Cognitive Science Society*, (Boston, MA).

Jones, M. N., and Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychol. Rev.* 114, 1–37.

Jonides, J., and Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychol. Bull.* 96, 29–44.

Khalkhali, S., Wammes, J., and McRae, K. (2012). Integrating words that refer to typical sequences of events. *Can. J. Exp. Psychol*. 66, 106–114.

Kwantes, P. J. (2005). Using context to build semantics. *Psychol. Bull. Rev.* 12, 703–710.

Landauer, T. K., and Dumais, S. T. (1994). "Latent semantic analysis and the measurement of knowledge," in *Educational Testing Service Conference on Natural Language Processing Techniques and Technology in Assessment and Education*, eds R. M. Kaplan and J. C. Burstein (Princeton, NJ: Educational Testing Service).

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240.

Levi, J. N. (1978). *The Syntax and Semantics of Complex Nominal*. New York, NY: Academic Press.

Lin, E. L., and Murphy, G. L. (2001). Thematic relations in adults' concepts. *J. Exp. Psychol. Gen.* 130, 3–28.

Lucas, M. (2000). Semantic priming without association: a meta-analytic review. *Psychon. Bull. Rev.* 7, 618–630.

Lynott, D., and Connell, L. (2010). Embodied conceptual combination. *Front. Psychol.* 1, 1–14. doi: 10.3389/fpsyg.2010.00212

Maki, W. S., and Buchanan, E. (2008). Latent structure in measures of associative, semantic, and thematic knowledge. *Psychon. Bull. Rev.* 15, 598–603.

Markman, A. B., and Wisniewski, E. J. (1997). Similar and different: the differentiation of basic level categories. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 54–70.

McKoon, G., and Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 1155–1172.

McRae, K., and Boisvert, S. (1998). Automatic semantic similarity priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 558–572.

McRae, K., Khalkhali, S., and Hare, M. (2012). "Semantic and associative relations in adolescents and young adults: examining a tenuous dichotomy," in *The Adolescent Brain: Learning, Reasoning, and Decision Making*, eds V. F. Reyna, S. Chapman, M. Dougherty, and J. Confrey (Washington, DC: APA), 39–66.

McRae, K., and Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so quickly. *Lang. Linguist. Compass* 3, 1417–1429.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., and Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *J. Mem. Lang.* 66, 545–567.

Mirman, D., and Graziano, K. M. (in press). Individual differences in strength of taxonomic versus thematic relations. *J. Exp. Psychol. Gen*. doi: 10.1037/a0026451. [Epub ahead of print].

Mirman, D., Walker, G. M., and Graziano, K. M. (2011). "A tale of two semantic systems: taxonomic and thematic knowledge," in *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, eds L. Carlson, C. Hoelscher, and T. F. Shipley (Austin, TX: Cognitive Science Society), 2211–2216.

Moss, H. E., Ostrin, R. K., Tyler, L. K., and Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: evidence from priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 863–883.

Murphy, G. L., and Brownell, H. H. (1985). Category differentiation in object recognition: typicality constraints on the basic category advantage. *J. Exp. Psychol. Learn. Mem. Cogn.* 11, 70–84.

Murphy, G. L., and Wisniewski, E. J. (2006). Familiarity and plausibility in conceptual combination: reply to Gagné and Spalding (2006). *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1438–1442.

Nation, K., and Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: evidence from semantic priming. *Cognition* 70, B1–B13.

Neely, J. H. (1991). "Semantic priming effects in visual word recognition: a selective review of current findings and theories," in *Basic Processes in Reading: Visual Word Recognition*, eds D. Besner and G. W. Humphreys (Hillsdale, NJ: Erlbaum), 264–336.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). *The University of South Florida Word Association, Rhyme, and Word Fragment Norms*. Available online at http://www.usf.edu/FreeAssociation/

Perraudin, S., and Mounoud, P. (2009). Contribution of the priming paradigm to the understanding of the coneptual developmental shift from 5 to 9 years of age. *Dev. Sci.* 12, 956–977.

Ratcliff, R., and McKoon, G. (1988). A retrieval theory of priming. *Psychol. Rev.* 95, 385–408.

Rosch, E. H. (1975). Cognitive representation of semantic categories. *J. Exp. Psychol.* 104, 192–233.

Rosch, E. H., and Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognit. Psychol.* 7, 573–605.

Sachs, O., Weis, S., Zellagui, N., Zvyagintsev, M., Mathiak, K., and Kircher, T. (2008). Automatic processing of semantic relations in fMRI: neural activation during semantic priming of taxonomic and thematic categories. *Brain Res.* 1218, 194–205.

Santos, A., Chaigneau, S. E., Simmons, W. K., and Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Lang. Cogn. Processes* 3, 83–119.

Sass, K., Sachs, O., Krach, S., and Kircher, T. (2009). Taxonomic and thematic categories: Neural correlates of categorization in an auditory-to-visual priming task using fMRI. *Brain Res.* 1270, 78–87.

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., Mirman, D., and Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8520–8524.

Shelton, J. R., and Martin, R. C. (1992). How semantic is automatic semantic priming? *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 1191–1210.

Simmons [Golonka], S., and Estes, Z. (2006). "Using latent semantic analysis to estimate similarity," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (Hillsdale, NJ: Lawrence Earlbaum Associates), 2169–2173.

Spalding, T. L., and Gagné, C. L. (2011). Relation priming in established compounds: facilitation? *Mem. Cogn.* 39, 1472–1486.

Spalding, T. L., and Ross, B. H. (2000). Concept learning and feature interpretation. *Mem. Cogn.* 28, 439–451.

Thompson-Schill, S. L., Kurtz, K. J., and Gabriele, J. D. E. (1998). Effects of semantic and associative relatedness on automatic priming. *J. Mem. Lang.* 38, 440–458.

Wisniewski, E. J. (1997). When concepts combine. *Psychon. Bull. Rev.* 4, 167–183.

Wisniewski, E. J., and Bassok, M. (1999). What makes a man similar to a tie? Stimulus compatibility with comparison and integration. *Cognit. Psychol.* 39, 208–238.

Wisniewski, E. J., and Love, B. C. (1998). Relations versus properties in conceptual combination. *J. Mem. Lang.* 38, 177–202.

Wisniewski, E. J., and Murphy, G. L. (2005). Frequency of relation type as a determinant of conceptual combination: a reanalysis. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 169–174.

Wu, L. L., and Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychol.* 132, 173–189.

Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 February 2012; accepted: 21 June 2012; published online: 11 July 2012.

*Citation: Jones LL and Golonka S (2012) Different influences on lexical priming for integrative, thematic, and taxonomic relations. Front. Hum. Neurosci. **6**:205. doi: 10.3389/fnhum.2012.00205*

APPENDIX A

Study 1 Stimuli.

Integrative Pairs

joint account	pest control	picnic lunch	executive secretary
head ache	nut cracker	copy machine	customer service
carbon atomv	computer disc	flea market	hair shampoo
snack attack	swan dive	jello mold	loan shark
data base	fashion fad	designer name	traffic sign
waste basket	football field	identification number	deposit slip
asteroid belt	mailbox flag	quart oil	laundry soap
cork board	antique furniture	cave opening	train station
gravy boat	net gain	dust pan	stick stone
ankle bracelet	herb garden	cuff pants	cross street
medicine cabinet	career goal	research paper	raspberry tart
birthday candle	camp ground	ice pick	property tax
health care	pool hall	nail polish	vodka tonic
sand castle	department head	will power	mountain top
brain cell	pledge honor	frog prince	bath towel
breakfast cereal	gray hound	bullet proof	personality trait
graduation ceremony	mint jelly	plum pudding	moral values
lawn chair	ski jump	equal rights	stair way
pipe cleaner	shop keeper	cardinal rule	cob web
comedy club	shoe lace	potato salad	trade wind
ethics code	sign language	tuna sandwich	
press conference	brick layer	skill saw	
mass confusion	shore line	split second	

Thematic Pairs

swamp alligator	couple date	bush leaves	sphinx pyramid
jungle animal	caravan desert	storm lightning	alligator river
medieval armor	physical doctor	vault lock	burglar robbery
fort army	evacuate earthquake	roach motel	tar roof
trombone band	siren emergency	rocks mountain	shovel sand
suds bath	headband exercise	physician needle	darkness scared
pillow blanket	home family	bat night	sailor sea
wedding bride	rooster farm	patient nurse	playground slide
herd buffalo	nun father	sponge ocean	kettle stove
desert cactus	sailing fishing	nurse patient	grief tears
cupboard cans	path forest	elephant peanut	cop ticket
blocks children	steak fries	fountain penny	bank vault
balloon clown	softball girls	sentence period	blood vein
sugar coffee	pasture grass	aircraft pilot	chef waiter
stadium concert	onion hamburger	buffalo plain	marriage wedding
crane construction	wheat hay	luggage plane	fireplace winter
cape coral	ape jungle	ace poker	giraffe zoo
accuse court	stove kitchen	algae pool	
haystack cow	otter lake	chapel priest	
spank cry	grass lawn	limousine prom	

(Continued)

Continued.

Taxonomic Pairs			
pronoun adjective	stag doe	branch leaf	broccoli spinach
lead aluminum	pound dollar	cougar lion	raccoon squirrel
wrist ankle	zebra donkey	shrimp lobster	veal steak
frustration anxiety	mongoose duck	yard mile	canal stream
handbag backpack	eye ear	compound mixture	lagoon swamp
jazz ballet	notebook folder	brother mother	bus taxi
baseball basketball	insight foresight	glands neck	rain thunder
cello bass	spatula fork	cousin nephew	muffin toast
bedroom bathroom	mink fox	nerve neuron	squash tomato
sofa bed	moss fungus	garlic onion	lip tongue
bathroom bedroom	colonel general	refrigerator oven	plaque trophy
champagne beer	brandy gin	drawing painting	soup vegetables
abdomen body	envy greed	carnival party	ceiling wall
broom brush	forehead hair	pear peach	south west
spinach cabbage	thigh hip	comet planet	rye wheat
turnip carrot	antler horn	metal plastic	cotton wool
ceramics crafts	cattle horse	bowl plate	editor writer
chalk crayon	earthquake hurricane	gallon quart	meter yard
cone cup	sweater jacket	mortgage rent	week year
son dad	swimmer jogger	pub restaurant	
moose deer	relish ketchup	plate saucer	
shelf desk	rope ladder	tenor soprano	

APPENDIX B

INSTRUCTIONS FOR RATING TASKS

Thematic relatedness

For each word pair (e.g., CREAM COFFEE; BLACKBOARD CHALK), you will be asked to rate the extent to which the concepts are linked together in a common scenario, event, or function. For example, CREAM is often added to COFFEE, and CHALK is used to write on a BLACKBOARD. Thus, these items are thematically connected to each other. Please note that thematically connected items may not necessarily share similar features. For example, BLACKBOARD and CHALK are different shapes and sizes and are made out of different materials. For each pair, rate the extent to which the pair is connected to some common theme (such as a classroom theme in the above example) on a scale from 1 (not thematically connected) to 7 (highly thematically connected). Please use the full range of the scale (1, 2, 3, 4, 5, 6, or 7) in indicating your responses.

Integrative ratings (sentence context)

You will judge the sensibility of a word pair (shown in ALL CAPS) within its sentence (e.g., George picked up the CEREAL BOWL or George picked up the PLATE BOWL). Please indicate your sensibility judgment on a scale from 1 (not at all sensible) to 7 (completely sensible). Please use the full range of the scale (1, 2, 3, 4, 5, 6, or 7) in indicating your responses.

Category relatedness ratings

You will rate the extent to which the two words (e.g., NECKLACE BRACELET, DUCK GOOSE) belong to the same specific category (e.g., BIRDS rather than the more general category ANIMALS). Please rate each word pair from 1 (not at all category co-members) to 7 (definitely members of the same specific category), and be sure to use the full range of the scale (enter 1, 2, 3, 4, 5, 6, or 7). For example, NECKLACE and BRACELET are both types of jewelry and thus would likely be given a high rating. In contrast, SILVER and BRACELET belong to different categories (SILVER is a type of metal, whereas BRACELET is a type of jewelry) and thus should be given a lower rating. Some items (HORSE CHICKEN) will belong to the same general category (HORSES and CHICKENS are both animals) but should also be given a lower rating as horses are a type of mammal and chickens are a type of bird.

Familiarity ratings

In the following experiment you will read a series of 195 word pairs (e.g., CREAM COFFEE; BLACKBOARD CHALK). For each word pair, please rate how unfamiliar or familiar the word pair is. For example, a word pair such as FRUIT BASKET might sound more familiar to you than the word pair DONKEY HILL. The scale ranges from 1 (unfamiliar) to 7 (familiar). Please use the full range of the scale (1, 2, 3, 4, 5, 6, or 7) in indicating your responses.

Feature similarity ratings

You will rate the similarity of the two words (e.g., DOTS STRIPES) on a scale from 1 (not at all similar) to 7 (very similar). Please use the full range of the scale (1, 2, 3, 4, 5, 6, or 7) in indicating your responses.

Two words are similar if they look alike or belong to the same category. For example, DOTS and STRIPES are similar (both are types of patterns or designs). However, SHIRT and STRIPES would not be similar. Even though stripes are often found on shirts, a shirt is a type of CLOTHING. Furthermore, whereas ZEBRA is associated with STRIPES, these two words are also not very similar, because they belong to different categories (i.e., animal and pattern categories).

APPENDIX C

Stimuli.

Target	Prime-type		
	Integrative	**Thematic**	**Taxonomic**
bacon	turkey	eggs	pork
book	instruction	editor	article
breakfast	hotel	bacon	lunch
cake	fruit	party	muffin
candy	chocolate	halloween	gum
carrot	garden	salad	beet
cat	alley	pet	lion
clock	antique	hour	watch
closet	broom	hanger	cabinet
coat	wool	lab	cape
cow	barn	dairy	pig
crown	jewel	queen	hat
doctor	animal	prescription	dentist
fries	carnival	steak	chips
guitar	strings	concert	drums
heart	donor	blood	liver
horse	trail	wagon	cow
hotdog	beef	mustard	sausage
jar	glass	jelly	bottle
lamp	street	bulb	flashlight
needle	steel	thimble	thorn
ocean	coral	shark	lake
organ	pipe	church	accordion
oyster	sea	pearl	scallop
panda	jungle	bamboo	grizzly
pants	linen	hem	suit
pencil	art	notebook	crayon
prison	inmate	guard	dungeon
rain	summer	hurricane	sleet
river	forest	boat	lake
robe	cotton	bath	cloak
saxophone	brass	jazz	clarinet
ship	battle	harbor	yacht
shirt	silk	tie	jacket
siren	ambulance	emergency	alarm
skill	job	expert	technique
skirt	suede	girl	shorts
smoke	signal	pollution	smog
soap	dishwasher	shower	shampoo
soccer	field	kick	volleyball
soup	tomato	bowl	chili
speech	history	campaign	lecture
spoon	silver	tea	knife
sport	contact	coach	tennis

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 19 October 2012
doi: 10.3389/fnhum.2012.00287

The cognitive chronometric architecture of reading aloud: semantic and lexical effects on naming onset and duration

Layla Gould[1], Jacqueline Cummine[2] and Ron Borowsky[1]*

[1] *Department of Psychology, University of Saskatoon, Saskatoon, SK, Canada*
[2] *Department of Speech Pathology and Audiology, University of Alberta, Edmonton, AB, Canada*

Edited by:
Penny M. Pexman, University of Calgary, Canada

Reviewed by:
Evan F. Risko, University of Memphis, USA
Michael Reynolds, Trent University, Canada

****Correspondence:***
Ron Borowsky, Department of Psychology, University of Saskatoon, Saskatoon, SK, Canada.
e-mail: ron.borowsky@usask.ca

We examined onset reaction time (RT) in a word naming task using an additive factors method (AFM). The pattern of additive and over-additive joint effects on RT among Instructions (INST: name all, name words), Word Frequency (WF: $\log_{10}$HAL), Semantic Neighborhood Density (SND: Inverse Ncount), and Word Type (WT: regular, exception) supported a cognitive chronometric architecture consisting of at least two cascaded stages of processing, with the orthographic lexical system as the locus of the INST × WF and the INST × SND interactions, and the phonological output system as the locus of the WF × WT and the SND × WT interactions. Additivity between INST and WT supports the notion that these variables affect separable systems, and a WF × SND interaction supports a common locus of their effects. These results support stage-like/cascaded processing models over parallel processing models of basic reading. We also examined response duration (RD) in these data by recording and hand-marking vocal responses, which provides evidence that basic reading processes are ongoing even after the initiation of a vocal response, and supports the notion that the more lexically a word is read, the shorter the RD. As such, the effects of WT and INST on RD were opposite to their effects on RT however the effects of WF and SND on RD were in the same direction as their effects on RT. Given the combination of consistent and dissociating effects between RT and RD, these results provide new challenges to all models of basic reading processes.

Keywords: reading aloud, semantic processing, lexical processing, naming response onset, naming response duration, word frequency, semantic neighborhood density, additive factors method

Semantic knowledge represents our worldly understanding of what things mean, how to interact with objects in our environment, how to interpret symbols and actions, as well as the meanings of words. As such, semantic knowledge is core to understanding not only language, but to understanding perception and cognition, and our world, in general. Although many years have been devoted to studying semantic knowledge, this concept has been a difficult one to elucidate due to its breadth. There are numerous ways to operationalize semantic processing, which provides multiple perspectives on the issue, but also broadens the problem space as opposed to narrowing it. However, as researchers have focused on and operationalized particular aspects of semantic knowledge, some substantial progress has been made (e.g., Balota et al., 2006; Yap et al., 2011).

Yap et al. (2012) recently demonstrated that semantic variables such as semantic neighborhood density (SND), number of features, semantic ambiguity (i.e., number of senses), imageability, and body-object interaction were reliable predictors of performance in several tasks of lexical processing. The only exceptions were the effects of SND and semantic ambiguity in the speeded pronunciation task. The null effect of semantic ambiguity in pronunciation has previously been argued to represent a lack of semantic influence in naming compared to lexical decision, for which there is an advantage for words with multiple meanings (Borowsky and Masson, 1996). Borowsky and Masson argued that the lexical decision task involves a monitoring of activation in orthographic, phonological, and semantic systems, thereby allowing for a familiarity-based lexical decision to benefit from multiple semantic representations (see also Balota and Chumbley, 1984; Chumbley and Balota, 1984), whereas naming can be accomplished without involvement of semantics and thus the lesser effect of semantic ambiguity in naming. It is possible that the effects of SND may behave similarly to the effects of semantic ambiguity, in that there may be an advantage for higher SND under conditions that encourage lexical access (see also Balota et al., 2004; Yap and Balota, 2009). One of the goals of the present research is to explore word naming behavior under conditions where lexical access is either compulsory or not. Another goal is to expand the investigation of naming behavior to more than just the onset of response, as has been done by Balota et al. (1989). Balota et al. explored duration of vocalizations in a semantic priming paradigm, similar to Balota and colleagues' work with other basic reading tasks involving parameters beyond response onset (e.g., Abrams and Balota, 1991; Bangert et al., 2012). As a general principle, going beyond the initial onset of response provides a larger window through which to view the effects of underlying cognitive processes. As perhaps the most ecologically valid basic reading task, the task of reading aloud is critical to explore in terms of both of our goals of manipulating lexical/semantic access and examining both response onset and duration.

METHODOLOGICAL CONSIDERATIONS IN READING ALOUD

The measurement of vocal onset *reaction time* (RT) has been central to research on basic cognitive processes since the invention of the voice-key (Dunlap, 1913; Boder, 1933). Although many researchers had initially assumed that the initiation of a vocal response first requires the generation of a complete phonological code for the entire word, this assumption has been challenged in recent years (e.g., Hudson and Bergman, 1985; cf. Rastle et al., 2000). Furthermore, research involving a delayed naming task (i.e., pronunciation is delayed until a cue is given) has demonstrated that the frequency effect still manifests in onset RT even after delays up to 1400 ms (Balota and Chumbley, 1985; see also Monsell et al., 1989). As such, it appears that the influences of lexical variables such as word frequency (WF) are still having an effect even after sufficient time to prepare and initiate a response. Delayed naming evidence notwithstanding, it is unclear why it would be necessary to hold off the initiation of the vocal response until the entire word is decoded, especially given the typical instructions (INST) to name words as quickly and accurately as possible. Furthermore, several models of reading refer to: (1) a relatively slow serial grapheme-to-phoneme translation system, which allows for the naming of novel words in a serial/cascaded fashion, as well as (2) a relatively fast lexical system, which allows words to be named in a "whole-word" manner (e.g., Coltheart et al., 2001; Coltheart, 2006; Borowsky et al., 2012). Nearly a century of research based on vocal onset RTs has been conducted to explore these and other basic reading processes. Given that cognitive processes could be operating beyond the initiation of vocal onset, it is important to explore measures of naming responses that go beyond measuring the onset. Another major goal of our present research involves exploring the *response duration* (RD) of vocal responses in addition to RT.

Research into the chronometric architecture of cognition also has a long history. Donders' (1969) subtractive logic provided the first method of examining when certain cognitive processes were occurring. For example, if one were to subtract the time that it takes to respond to the presence or absence of a flash of light, from the time that it takes to respond to a flash of light of a certain color, one could attribute the difference in time to color processing. However, this logic requires the untenable assumption of *pure insertion*, whereby more than just color processing has been inserted into the task (e.g., holding in memory the instructed target color). Sternberg (1969) argued that *pure insertion* was not a tenable assumption, and developed the Additive Factors Method (AFM) (see also: Borowsky and Besner, 1993; Roberts and Sternberg, 1993; Stolz and Neely, 1995; Yap and Balota, 2007). By looking at the joint effects of the variables, this method allows for the examination of whether two variables are affecting the same system in time (i.e., over-additive interactive effects on RT) or separable systems in time (i.e., additive effects on RT). Another major goal of our present research involves exploring the joint effects of four variables that are known to reflect the operation of subsystems of basic reading processes: SND, WF, WT, and INST.

EFFECTS THAT REFLECT THE OPERATION OF LEXICAL/SEMANTIC SUBSYSTEMS

As described earlier, semantic knowledge is core to any model of language processing. SND has been shown to be a measure of semantic processing (Shaoul and Westbury, 2010). This measure reflects the number of words that co-occurred with a target word within a fixed distance threshold, as determined by an analysis of 57,153 words present in Wikipedia in April 2010 (a total of 971,819,808 occurrences). Words that have a large number of semantic neighbors show benefits relative to words that have a small number of semantic neighbors, as was shown by Yap et al. (2012) using the tasks of: lexical decision, go/no-go lexical decision, speeded naming, progressive demasking, and semantic classification. SND could serve to facilitate semantic processing, as well as connections to other word-level systems such as the orthographic lexical system and the phonological output system, in that the higher the SND the higher the number of facilitative connections both within and between levels (as is typical of interactive activation architectures, McClelland and Rumelhart, 1981; Coltheart, 2006; Yap et al., 2012; see **Figure 1**).

There are several models that propose that printed WF effects manifest in the lexical/semantic systems (e.g., Morton, 1979; McClelland and Rumelhart, 1981; McCann and Besner, 1987; Borowsky and Besner, 1993; Reynolds and Besner, 2005). For example, WF could affect the connections between the lexical subsystems and semantic system, whereby the more frequently a word is read, the faster the rate of activation in these systems, and the faster the RT (see **Figure 1**; Borowsky and Besner, 1993). The WT [i.e., regular vs. exception words (EXCs)] effect on RT is another effect that reflects basic reading processes. Given that regular words (REGs) can be pronounced correctly through both the sublexical grapheme-to-phoneme conversion (GPC) route (allowing the word to be "sounded out") or the orthographic lexical route (allowing the word to be read in a "whole-word" manner), these routes produce the same pronunciation at the phonological output system. Conversely, EXCs must be processed via the orthographic lexical route to be pronounced correctly. EXCs produce a slower RT because the two routes produce conflicting pronunciations, and a single phonological output must ultimately be selected, particularly in the case of low frequency EXCs.

WT has also been found to interact with WF on naming RT, whereby EXCs produce slower RTs and elicit a greater WF effect, compared to REGs (e.g., Monsell et al., 1992; Cummine et al., 2010). This same interactive pattern has been shown on the Blood Oxygenation Level Dependent (BOLD) function intensity in the supplementary motor area (SMA), which likely represents the phonological output system given that the SMA is the last cortical region prior to activating the motor cortex (see **Figure 1**; Cummine et al., 2010). Other reading route reliance effects have also been reported, whereby there is flexibility on route reliance depending on stimulus and task manipulations (e.g., Rastle and Coltheart, 1999; Zevin and Balota, 2000; Borowsky et al., 2002; Reynolds and Besner, 2005; see Balota et al., 2006 for a review).

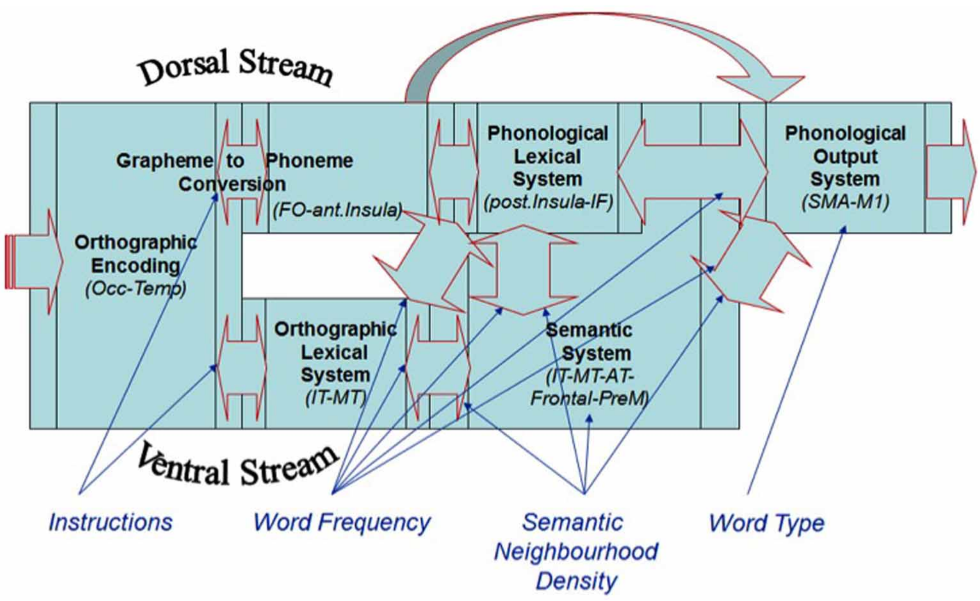


FIGURE 1 | A dual-stream, ventral-lexical, dorsal-sublexical, cascaded processing framework for basic reading processes.

Given the proximity of SND effects to WF effects in the basic reading architecture, in that they both involve lexical/semantic systems, it seems reasonable that SND and WF should also interact due to these common influences.

Researchers have begun to explore the strategic effects of INST on reading. For example, Hino and Lupker (2000, see also Kinoshita and Woollams, 2002) presented participants with a list of words and non-words (NWs), and used what we refer to as a *name words* condition and a *name all* condition. INST to *name words* required the participant to name a stimulus aloud only if it spells a word, which forces the participant to process the word via the orthographic lexical route, as they must first verify that the stimulus is in fact a word (see **Figure 1**). Cummine et al. (2012) provided direct functional and behavioral evidence that INST to *name words* forces reliance on the orthographic-lexical route. We reported that INST to *name words* showed greater visible activation along the ventral-lexical stream, as well as produced greater WF effects on RT relative to INST to *name all* stimuli. There was no interaction between INST and WT (i.e., additivity), but there was an interaction between WF and WT under the normal *name all* instruction condition. When the AFM is employed, additive and interactive joint effects can reveal the loci of effects among the processing systems, and how many systems are involved in the cognitive chronometric architecture.

ADDITIVE FACTORS METHOD

The AFM proposes that if two variables interact over-additively on RT (such as the WF × WT interaction described above), it is indicative of those variables affecting a common system of processing in time (see **Figure 2**). Also, the over-additive interaction between INST and WF is indicative of these two variables affecting a common system (Cummine et al., 2012). In contrast, if two variables produce additive effects on RT, those variables are assumed to be affecting separable (even if they are cascaded; McClelland, 1979) systems of processing (see **Figure 3**). As such, the additive pattern found between INST and WT is taken to indicate that those variables are affecting separable systems of processing. Taken together, these joint effects support a cognitive chronometric architecture of at least two systems (see **Figure 4**), whereby INST and WF interact in a relatively early system that serves to gate the processing of words through the orthographic lexical route when the INST are to *name words* only (resulting in a

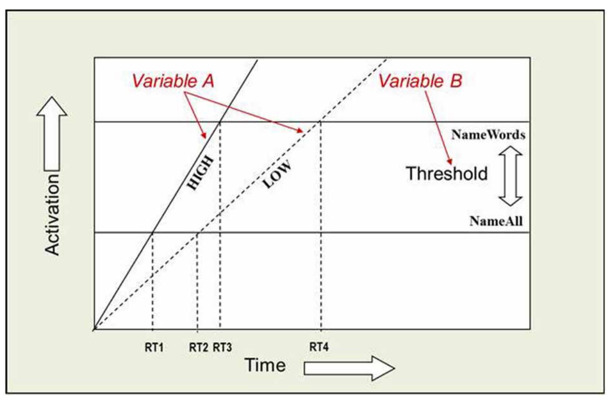


FIGURE 2 | Over-additive interaction between INST and WF. An Additive Factors Method interpretation of this interaction is that both INST and WF are affecting a common system in time. If INST are assumed to affect the threshold for activation, and WF is assumed to affect the rate of activation over time, or vice versa, then the points in time when each rate crosses a threshold correspond to the average onset RTs. [(RT4 − RT2) > (RT3 − RT1)] and [((RT2 + RT4)/2) > ((RT1 + RT3)/2)]. WF and WT interact in a similar fashion.

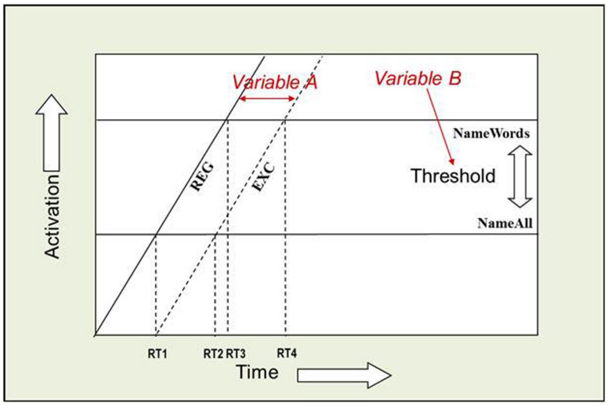


FIGURE 3 | Additive joint effects between INST and WT. An Additive Factors Method interpretation of this additive effect is that INST and WT are affecting separable systems in time. If INST are assumed to affect the threshold for activation (i.e., the amount of time it takes to verify that a letter string spells a word), and WT is assumed to shift the rate of activation over time (i.e., the time it takes to choose among the competing phonological codes for EXCs), or vice versa, then the points in time when each rate crosses a threshold correspond to the average onset RTs. [(RT4 − RT2) = (RT3 − RT1)] and [((RT2 + RT4)/2) > ((RT1 + RT3)/2)].

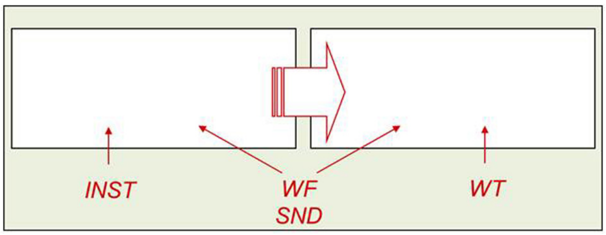


FIGURE 4 | A cognitive chronometric architecture to account for the INST × WF, INST × SND, WF × WT, and SND × WT over-additive interactions, and INST + WT additivity.

lower threshold, see **Figures 2** and **3**), and WF and WT interact in the relatively late phonological output system in a similar fashion. The present research examines SND in addition to these variables in order to further constrain the architecture for basic reading processes. Given that WF and SND affect some common systems and connections, one would expect that SND should show similar joint effects with INST and WT as did WF (see **Figures 1** and **4**).

NAMING RESPONSE DURATION

As previously mentioned, one major goal of our present research involves not only exploring measures of onset RT, but also the RD of vocalizations. Previous research on naming responses has almost solely relied on voice-key measures of *onset* RT. However, given that basic reading processes may still be operating while initiating a vocal response, onset RT measures may not be comprehensive enough. Examining the *duration* of vocal responses should serve as an important additional measure of word processing. Balota et al. (1989) showed that RD is an important converging measure of reading processing in that naming RDs are significantly shorter when a related word is presented. Given that their relatedness manipulation served to decrease RD, this supports the notion that manipulations that enhance lexical/semantic processing yield shorter RDs. Thus, one can predict that shorter RDs should reflect lexical "whole-word" reading, whereas longer RDs should reflect sublexical GPCs. Consistent with the view that RDs can reflect important aspects of cognitive processing post-onset RT, Kawamoto et al. (1999) research on onset and rime durations suggests that the criterion to initiate pronunciation is based on the initial phoneme and not the whole-word. That said, RD effects could also reflect how familiar we are with a given word's pronunciation, whereby the more often we pronounce a given word, the shorter the RD could get as a function of the word being read more lexically over time.

Our present research also contributes a novel means of measuring RT and RD in word recognition, whereby one manually analyzes speech envelopes of verbal responses to objectively measure the onset and offset of a naming response. Previous studies have found that using voice-keys to measure onset RT may be quite unreliable. For example, a study by Rastle and Davis (2002) found that different types of voice-keys can produce different results, whereby the voice-keys were often triggered at different points in time following the actual onset of the naming response. They found that hand-marking the acoustic onset of each word using visual waveforms of intensity over time can produce less error than voice-keys, and thus suggest that visually investigating these waveforms may be an important way to check voice-key onsets. We also note that the proportion of errors that are due to voice-key problems can be quite substantial (e.g., Balota et al.'s, 2007, large scale study of naming reported that nearly 13% of naming errors were due to voice-key problems). Our labs has begun to digitally record participants' vocal responses and then manually inspect each vocal response using PRAAT digital software (Boersma and Weenink, 2012). By using both visual and auditory cues to identify vocalization onset and offset, it allows us to precisely measure the onset RT and RD of each response. Given the potential for variability of voice-key measurements, in the present experiments we analyze the RTs obtained via hand-marking (Experiment 1), and RTs obtained via a voice-key (Experiment 2). Importantly, we also measure the RDs via hand-marking in both experiments, which has not been previously reported.

HYPOTHESES

Our first set of hypotheses involves the joint effects of INST, WF, SND, and WT on onset RT (see **Figure 1**). We predict that: (1) *INST × WF*—to the extent that INST and WF are both affecting the orthographic lexical system, they should show an over-additive interaction on onset RT; (2) *INST + WT*—because we are predicting that INST are having their effect early by gating processing towards the orthographic lexical system under *name words* INST, whereas WT has been shown to have its effect later at the phonological output system, these two variables should show an additive pattern on RT; (3) *INST × SND*—to the extent that INST and SND are both affecting the orthographic lexical system, they should show an over-additive interaction on onset RT; (4) *WF × SND*—to the extent that WF and SND are affecting the semantic system and connections to other word-level systems, they should also show an over-additive interaction on onset RT; (5) *SND* × *WT*—to the extent that SND and WT are affecting the phonological output system, they should also show an over-additive interaction on onset RT; and (6) *WF* × *WT*—to the extent that WF and WT are affecting the phonological output system, they should produce the typical over-additive interaction under normal *name all* INST.

Our second set of hypotheses are in regards to the RDs of vocal responses: *EXC RD* < *REG RD* < *NW RD*—given that EXCs must be processed as whole-words and read via the relatively fast lexical system in order to be pronounced correctly, they should produce the shortest RDs, despite the fact that EXC onset RTs are longer than REG onset RTs. Given that NWs must be processed through the relatively slow sublexical GPC system, they should produce the longest RDs. Finally, given that REGs can be processed through either route, they should elicit intermediate RDs relative to EXCs and NWs, despite having the fastest RTs[1]. With respect to INST, given that the *name words* condition also forces participants to rely on the orthographic lexical system, it should produce shorter RDs compared to the *name all* condition (*name words RD* < *name all RD*). Given Balota et al.'s demonstration of semantic priming having a facilitative effect on naming onset RT and naming RD, the SND effect should remain in the same facilitative direction for onset RT and RD in the present experiment. Furthermore, given that WF is considered to have its effects at the same semantic/lexical level as SND, the WF effect should also remain in the same facilitative direction for onset RT and RD.

In the experiments that follow, Experiment 1 ($n = 20$) was conducted in a MRI scanner (see Cummine et al., 2012), and Experiment 2 ($n = 40$) was conducted in a behavioral lab. Although the results of these experiments are presented separately, we will focus our discussion on analyses that combine the data from Experiments 1 and 2.

EXPERIMENT 1

METHODS

Participants

Twenty participants responded to a local advertisement for a fMRI experiment at the University of Alberta (see Cummine et al., 2012 for details). The experiment was performed in compliance with the relevant laws and institutional guidelines, and was approved by the University of Alberta Health Research Ethics Board. The participants' consent was obtained according to the Declaration of Helsinki (1996). Inclusion criteria consisted of normal or corrected-to-normal vision and English as a first language. Eighteen participants were right-handed.

Stimuli

One hundred and twenty-six pairs of monosyllabic REGs and EXCs matched for initial onset and length were used as critical stimuli (Patterson and Hodges, 1992). SND, as described earlier in the introduction, was measured using inverse Ncount (Shaoul and Westbury, 2010), which is the inverse of the number of semantic neighbors +1. These stimuli were well-matched on several of the characteristics available from the E-Lexicon Database (http://elexicon.wustl.edu/, Balota et al., 2007), as we found that WT (REG = 0, EXC = 1) did not correlate significantly with $\log_{10}$ HAL WF ($r = 0.036$, $p = 0.57$), bigram frequency by position ($r = 0.033$, $p = 0.60$), bigram mean frequency, ($r = -0.051$, $p = 0.42$), bigram sum frequency ($r = -0.036$, $p = 0.57$), number of morphemes ($r = 0.058$, $p = 0.357$), number of phonemes ($r = -0.082$, $p = 0.20$), phonological neighborhood ($r = 0.081$, $p = 0.202$), or inverse Ncount ($r = -0.031$, $p = 0.50$). These words can be considered to be of fairly high familiarity, as their mean WF is relatively high ($\log_{10}$ HAL WF mean = 9.63). A set of 128 pronounceable NWs were also generated from the critical words by changing one or two letters. The mean length of the NWs (4.48 letters) was well matched to the mean length of the words (4.51 letters for both the EXCs and REGs), [$t_{(252)} = 0.307$, $p = 0.759$]. For each INST condition, a total of 190 stimuli were presented in two blocks (one block had 31 REGs, 32 EXCs, and 32 NWs, and the other block had 32 REGs, 31 EXCs, and 32 NWs), such that every participant was presented with each stimulus only once, and stimuli were cycled through INST conditions across participants so that each stimulus was presented equally often under each INST set.

Procedure and apparatus

For the *name all* INST condition, participants were instructed to "read aloud each letter string, as quickly and accurately as possible." For the *name words* INST condition, participants were instructed to "only read aloud each letter string that spells a word, as quickly and accurately as possible." Letter strings were presented, and participants responded vocally, during a regular periodic gap in the image acquisition that followed the offset of each volume of images (i.e., a sparse-sampling, or gap, paradigm; Borowsky et al., 2005a,b, 2006, 2007, 2012; Cummine et al., 2010, 2012). That is, a letter string was presented for 1850 ms during a silent gap, at the offset of a 1850 ms acquisition of a volume of images, allowing participants to name aloud the letter string immediately and without gradient noise in the background. Letter strings were randomly selected, without repetition, and back-projected one at a time on a screen such that they were visible to the participants through the mirror on the head coil. A computer running EPrime software (Psychology Software Tools, Inc., http://www.pstnet.com) was used to trigger each image acquisition in synchrony with the presentation of visual stimuli.

Vocal responses were recorded at 96 KHz, 24 bit, through the intercom using an Olympus LS11 digital recorder, during the acquisition gap. These recordings were then analyzed using PRAAT software (Boersma and Weenink, 2012), and the speech spectrograms and formants were used to localize vocalization onset RT and the RD. Given that the gradient noise associated with the final image acquisition in each volume coincided with

[1] Following the interpretation of the WF × WT interaction within a dual-route framework (e.g., Coltheart et al., 2001) the initiation of a response is modulated by the consistency in the computed phonological codes from sublexical and lexical routes. On one hand, in the case of EXCs, while the onset RT may be delayed due to the competing phonological codes, the RD should be relatively short given that there is greater reliance on the lexical system. Regular words, on the other hand, have a relatively fast onset RT because there is no competition from the sublexical and lexical routes. However, information from both the sublexical and lexical systems can contribute to correct responding and thus the RD may be longer to accommodate the inclusion of the sublexical information.

the onset of the target stimulus, we were able to use it as an auditory and visual cue on the digital recording for identifying when the stimulus appeared on the screen (see **Figure 5**). By replaying the audio recording, we were able to code whether each participant's response was correct, incorrect, or a spoiled trial. By using PRAAT to analyze the speech spectrograms, and by replaying the audio recording, we were able to determine the exact time point for the onset RT and the RD.

RESULTS

Word naming reaction time

The naming onset RT data were first aggregated by participant as a function of INST (*name all, name words*) and WT (REG and EXC). Medians of the correctly named item RTs were submitted to a 2 × 2 general linear model (GLM) ANOVA, with WT and INST as repeated measures factors. The median naming onset RTs are presented in **Figure 6**. There were significant main effects of INST, [$F_{(1,\ 19)} = 7.626$, $MSe = 4659$, $p = 0.01$], and WT, [$F_{(1,\ 19)} = 4.97$, $MSe = 239$, $p = 0.04$], and no significant interaction, [$F_{(1,\ 19)} = 0.049$, $MSe = 442$, $p = 0.83$].

Non-word naming reaction time

The NW condition in the *name all* INST condition yielded a median onset RT of 687.7 ms (Loftus and Masson, 1994, repeated-measure 95% confidence interval (CI) = ±21.9).

Accuracy

The mean accuracy rates resulted in 100% accuracy in all cells, and thus there was no variance for a statistical analysis.

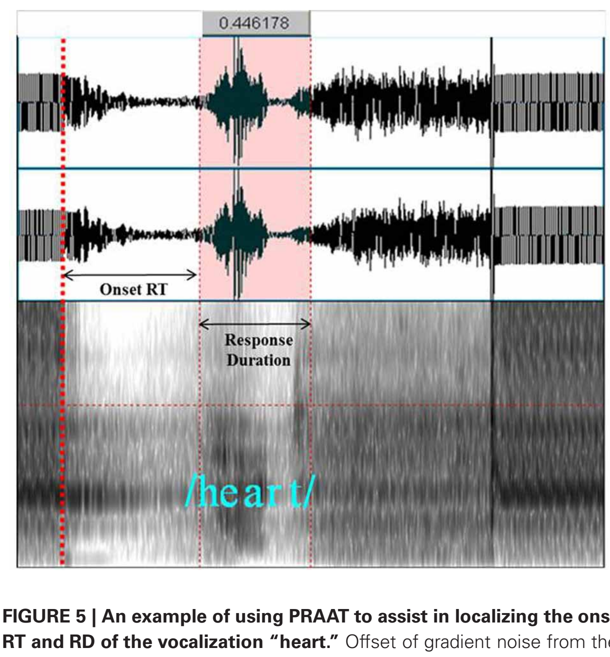


FIGURE 5 | An example of using PRAAT to assist in localizing the onset RT and RD of the vocalization "heart." Offset of gradient noise from the MRI can be seen and heard at the time point of the coarse red-dashed line (relevant for Experiment 1 only). The onset and offset of the vocalization can be seen and heard between the thin red-dashed lines, while the temporal distance between those lines (i.e., the RD) is indicated at the top. Visual inspection of both the spectrograms and the formants, as well as several replayings of the audio recording, allowed for precise measurement of the onset RT and RD.

Word frequency effects on reaction time

In order to evaluate the effects of WF as a continuous variable, GLM regressions were conducted on each participant's correct onset RTs, with RT as the dependent variable, and WF as a continuous independent variable, separately for each combination of INST and WT. The resulting WF coefficients for each INST and WT set were then aggregated over participants (e.g., Borowsky et al., 2002), and submitted to a 2 × 2 GLM ANOVA, with WT and INST as repeated measures factors. This analysis allows one to generalize to both items and participants, in that items are treated as the unit of analysis in the regressions, and that participants are treated as the unit of analysis when the co-efficients are being statistically tested. Given our use of the AFM and a focus on two-way joint effects, interaction effects were restricted to two-way joint effects in all the analyses reported here. **Figure 6** shows the mean co-efficients above the median RTs. There was a significant main effect of INST on the size of WF effect, [$F_{(1,\ 19)} = 15.5$, $MSe = 83.6$, $p = 0.001$], which represents an INST by WF interaction on naming RT, whereby the WF effects are greater under *name words* INST. The main effect of WT on the size of WF effect was not significant, [$F_{(1,\ 19)} = 1.38$, $MSe = 115.2$, $p = 0.25$]. An analysis of this WF by WT interaction for *name all* INST failed to show a significant effect, [$t_{(19)} = -1.27$, $SEM = 3.47$, $p = 0.22$].

Semantic neighborhood density effects on reaction time

In order to evaluate the effects of SND as a continuous variable, GLM regressions were conducted on each participant's correct onset RTs, with RT as the dependent variable, and SND as a continuous independent variable, separately for each combination of INST and WT. The resulting SND co-efficients for each INST and

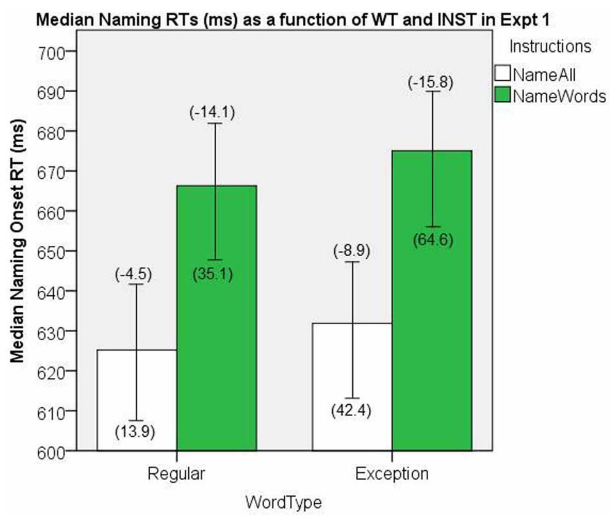


FIGURE 6 | Median Naming RTs (in ms) as a function of WT and INST in Experiment 1. The 95% C.I.s are presented as error bars using Loftus and Masson's (1994) method. Co-efficients relating WF to RT (ms/$\log_{10}$HALWF) are presented in parentheses above each error bar, and co-efficients relating SND to RT (ms/unit inverseNcount) are presented in parentheses below each error bar.

WT set were then aggregated over participants, and submitted to a 2 × 2 GLM ANOVA, with WT and INST as repeated measures factors. **Figure 6** shows the mean co-efficients below the median RTs. There was a significant main effect of INST on the size of the SND effect, [$F_{(1,\ 19)} = 5.53$, $MSe = 1698.7$, $p = 0.03$], which represents an INST by SND over-additive interaction on naming RT, whereby the SND effects are greater under *name words* INST. There was also a significant main effect of WT on the size of the SND effect, [$F_{(1,\ 19)} = 8.69$, $MSe = 1938.8$, $p = 0.008$], which represents a WT by SND over-additive interaction on naming RT, indicating that the SND effect is greater for EXCs than REGs.

Word naming response duration

The naming RD data were aggregated by participant as a function of INST and WT. Medians of the correctly named item RDs were submitted to a 2 × 2 GLM ANOVA, with WT and INST as repeated measures factors. The median naming RDs are presented in **Figure 7**. There was a significant main effect of WT, [$F_{(1,\ 19)} = 152.06$, $MSe = 90.3$, $p < 0.001$]. The main effect of INST showed a trend in the predicted direction but was not significant, [$F_{(1,\ 19)} = 1.72$, $MSe = 749.25$, $p = 0.20$], and no significant interaction, [$F_{(1,\ 19)} = 0.710$, $MSe = 105.7$, $p = 0.41$].

Non-word naming response duration

The NW condition in the *name all* INST condition yielded a median RD of 587.9 ms (Loftus and Masson, 1994, repeated-measure 95% CI = ±10.78).

Word frequency effects on response duration

We conducted analyses of WF effects on RD in the same way as our analyses on RT. **Figure 7** shows the mean co-efficients above the median RDs. There was a significant main effect of WT on the size of WF effect, [$F_{(1,\ 19)} = 9.42$, $MSe = 16.1$, $p = 0.006$], which represents a WT by WF interaction on naming RD, whereby the WF effects are greater for REGs. A main effect of INST on the size of WF effect approached significance, [$F_{(1,\ 19)} = 3.67$, $MSe = 17.3$, $p = 0.07$], whereby there was a tendency for an interaction, such that there were larger WF effects in the *name all* condition.

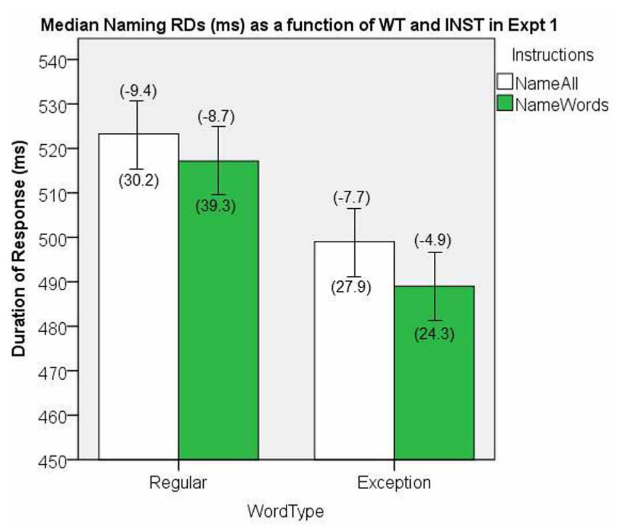


FIGURE 7 | Median Naming RDs (in ms) as a function of WT and INST in Experiment 1. The 95% C.I.s are presented as error bars using Loftus and Masson's (1994) method. Co-efficients relating WF to RT (ms/$\log_{10}$HALWF) are presented in parentheses above each error bar, and co-efficients relating SND to RT (ms/unit inverseNcount) are presented in parentheses below each error bar.

Semantic neighborhood density effects on response duration

We conducted analyses of SND effects on RD in the same way as our analyses on RT, and the mean co-efficients are shown below the median RDs in **Figure 7**. There was no significant main effect of INST on the size of the SND effect, [$F_{(1,\ 19)} < 1$, $MSe = 706.2$, $p = 0.99$], nor was there a significant main effect of WT on the size of the SND effect, [$F_{(1,\ 19)} = 2.15$, $MSe = 880.0$, $p = 0.16$], which suggests there were no interactions between INST and SND, or WT and SND on RD.

DISCUSSION

For onset RT there was a significant main effect of WT and INST, but no interaction. This additive pattern supports the notion of WT and INST affecting separable systems (see **Figure 4**). The onset RT analysis involving WF supports an over-additive interaction with INST. This pattern of interaction with WF supports the notion of INST affecting the same system as that affected by WF. Our analysis of the SND effect on onset RT showed an over-additive interaction between INST and SND, as well as between SND and WT. This pattern of interaction with SND supports the notion of INST and WT both affecting the same system as that affected by SND.

In keeping with our hypotheses regarding the RDs of vocal responses, the pattern of results supported: *EXC RD* < *REG RD* < *NW RD*—in that the main effect of WT was significant, and that the 95% CI for NWs did not overlap with any of the comparison cells. Furthermore, there was a trend for the *name words* INST condition to have shorter RDs than the *name all* INST condition.

Our analysis of the WF effect on RD revealed a very interesting pattern. Specifically, larger WF effects are associated with the longer RD cells (i.e., REGs), despite the fact that the opposite pattern was demonstrated for onset RT. As such, RD is shorter for the lexically read EXCs, which supports our hypotheses about shorter RDs being associated with lexically read items. Our analysis of the SND effect on RD showed no significant Two-Way interactions.

EXPERIMENT 2

METHODS

Participants

Forty undergraduate students participated for course credit in their introductory psychology class. The experiment was performed in compliance with the relevant laws and institutional guidelines, and was approved by the University of Saskatchewan Research Ethics Board. Inclusion criteria consisted of normal or corrected-to-normal vision and fluency in English. Thirty eight participants were right-handed.

Stimuli, procedure, and apparatus

These were identical to Experiment 1, with the following exceptions: Testing was done in a sound attenuated behavioral lab, stimuli were presented on a 15" Samsung CRT monitor connected to a PC running EPrime software, participants initiated each trial by pressing a button on the PST serial response box, which was also connected to a microphone that was interfaced with the voice-key for detecting vocalization onsets. RD was measured in the same way as Experiment 1 (see **Figure 5**).

RESULTS

The same analyses were conducted as in Experiment 1. One participant elicited error rates that were in excess of three SDs below the mean of all participants, and was thus excluded from the analyses.

Word naming reaction time

The median naming onset RTs are presented in **Figure 8**. There were significant main effects of INST, [$F_{(1, 38)} = 26.5$, $MSe = 2260.5$, $p < 0.001$], and WT, [$F_{(1, 38)} = 4.96$, $MSe = 686$, $p = 0.03$], and no significant interaction, [$F_{(1, 38)} = 0.463$, $MSe = 242$, $p = 0.96$].

Non-word naming reaction time

The NW condition in the *name all* INST condition yielded a median onset RT of 750.8 ms (Loftus and Masson, 1994, repeated-measure 95% CI = ±14.7).

Accuracy

The mean proportion accuracy rates are presented in **Figure 9**. There was a significant main effect of INST, [$F_{(1, 38)} = 15.62$, $MSe = 0.003$, $p < 0.001$], and WT, [$F_{(1, 38)} = 11.44$, $MSe = 0.001$, $p = 0.002$], and there was no significant interaction, [$F_{(1, 38)} = 2.64$, $MSe = 0.001$, $p = 0.11$]. The NW accuracy in the *name all* INST condition yielded a mean proportion of 0.90 (Loftus and Masson, 1994, repeated-measure 95% CI = ±.017).

Word frequency effects on reaction time

We conducted analyses of WF effects on RT in the same way as our analyses in Experiment 1. **Figure 8** shows the mean co-efficients above the median RTs. There was a significant main effect of INST on the size of the WF effect, [$F_{(1, 38)} = 11.28$, $MSe = 154.7$, $p = 0.002$], which represents a INST by WF interaction on naming RT, whereby the WF effects are greater for *name words* INST. The main effect of WT on the size of the WF effect (i.e., the WF by WT interaction) did not reach significance, [$F_{(1, 38)} = 1.78$, $MSe = 114.1$, $p = 0.19$], however, an analysis of this WF by WT interaction for the normal *name all* INST showed a significant effect, [$t_{(38)} = -2.13$, $SEM = 2.30$, $p = 0.04$].

Semantic neighborhood density effects on reaction time

We conducted analyses of SND effects on RT in the same way as our analyses in Experiment 1. **Figure 8** shows the mean co-efficients below the median RTs. There was a significant main effect of INST on the size of the SND effect, [$F_{(1, 38)} = 5.67$, $MSe = 3933.8$, $p = 0.02$], which represents an INST × SND over-additive interaction, whereby the SND effect is larger for the *name words* INST condition. There was also a significant main effect of WT on the size of the SND effect, [$F_{(1, 38)} = 4.46$, $MSe = 5534.5$, $p = 0.04$], which represents a SND × WT over-additive interaction, whereby the SND effect is larger for EXCs than for REGs.

Word naming response duration

The median naming RDs are presented in **Figure 10**. There was a significant main effect of WT, [$F_{(1, 38)} = 140.29$, $MSe = 210.2$,

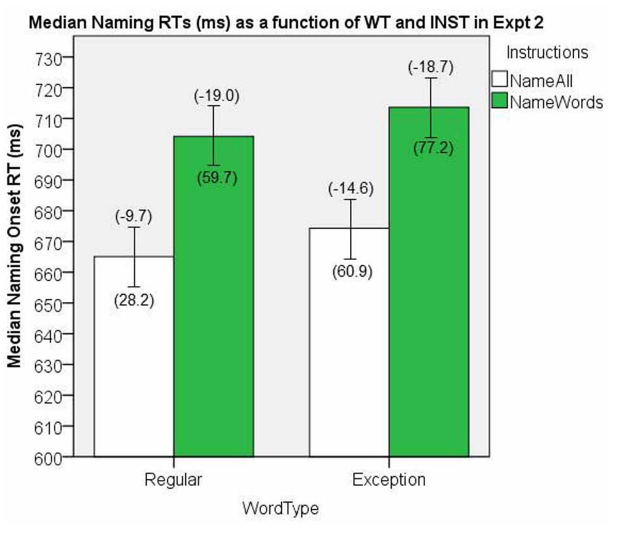


FIGURE 8 | Median Naming RTs (in ms) as a function of WT and INST in Experiment 2. The 95% C.I.s are presented as error bars using Loftus and Masson's (1994) method. Co-efficients relating WF to RT (ms/$\log_{10}$HALWF) are presented in parentheses above each error bar, and co-efficients relating SND to RT (ms/unit inverseNcount) are presented in parentheses below each error bar.

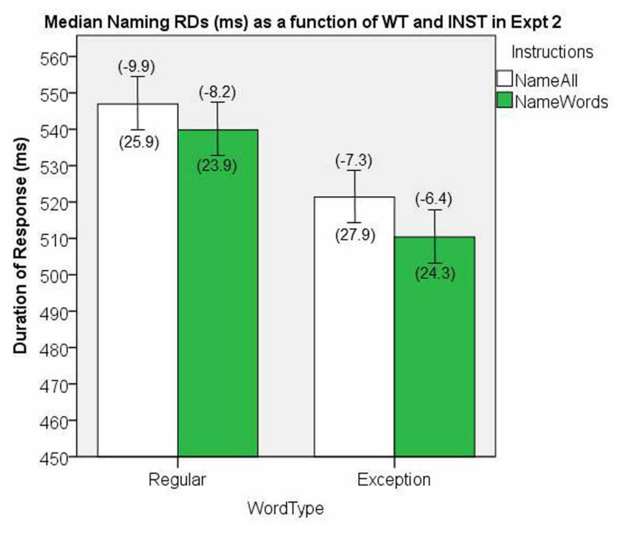


FIGURE 9 | Mean Proportion Accurate as a function of WT and INST in Experiment 2. The 95% C.I.s are presented as error bars using Loftus and Masson's (1994) method.

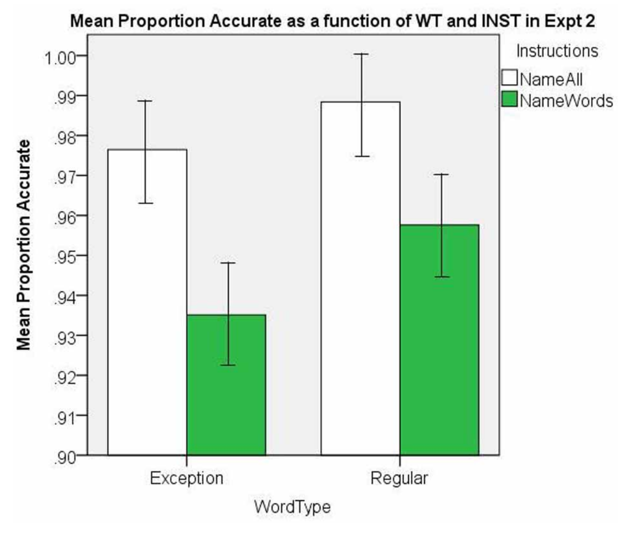


FIGURE 10 | Median Naming RDs (in ms) as a function of WT and INST in Experiment 2. The 95% C.I.s are presented as error bars using Loftus and Masson's (1994) method. Co-efficients relating WF to RT (ms/$\log_{10}$HALWF) are presented in parentheses above each error bar, and co-efficients relating SND to RT (ms/unit inverseNcount) are presented in parentheses below each error bar.

$p < 0.001$], which was in the predicted direction with EXCs showing shorter RDs. The main effect of INST approached significance, [$F_{(1, 38)} = 2.7$, $MSe = 1187.8$, $p = 0.10$], which was also in the predicted direction, and there was no significant interaction, [$F_{(1, 38)} = 0.766$, $MSe = 186.3$, $p = 0.39$].

Non-word naming response duration

The NW condition in the *name all* INST condition yielded a median RD of 561.5 ms (Loftus and Masson, 1994, repeated-measure 95% CI = ±7.65).

Word frequency effects on response duration

We conducted analyses of WF effects on RD in the same way as our analyses in Experiment 1. **Figure 10** shows the mean co-efficients above the median RDs. There was a significant main effect of WT on the size of the WF effect, [$F_{(1, 38)} = 4.28$, $MSe = 45.2$, $p = 0.045$], which represents a WT by WF interaction on naming RD, whereby the WF effects are greater for REGs. There was no main effect of INST on the size of WF effect, [$F_{(1, 38)} = 1.42$, $MSe = 44.2$, $p = 0.24$], although we note that the direction of the effects was consistent with Experiment 1.

Semantic neighborhood density effects on response duration

We conducted analyses of SND effects on RD in the same way as our analyses in Experiment 1. **Figure 10** shows the mean co-efficients below the median RDs. There was no significant main effect of INST on the size of the SND effect, [$F_{(1, 38)} = 0.17$, $MSe = 1826.0$, $p = 0.69$], nor was there a significant main effect of WT on the size of the SND effect, [$F_{(1, 38)} = 0.08$, $MSe = 652.1$, $p = 0.78$].

DISCUSSION

Our first set of hypotheses involved the joint effects of INST, WF, SND, and WT on onset RT. Consistently in both Experiments 1 and 2, we showed that: (1) *INST* × *WF*—the over-additive INST × WF interaction was significant, supporting the notion that these variables are affecting the orthographic lexical system; (2) *INST* + *WT*—these two variables showed an additive pattern on RT, supporting the notion that they are affecting separable systems, namely the orthographic lexical system and the phonological output system, respectively; (3) *INST* × *SND*—the over-additive INST × SND interaction was significant, supporting the idea that the orthographic lexical system is affected by both variables; (4) *SND* × *WT*—the over-additive SND × WT interaction was also significant, which is congruent with SND and WT both affecting the phonological output system; (5) *WF* × *WT*—the WF × WT over-additive interaction under the normal *name all* INST was significant, supporting the notion that these variables affect the phonological output system.

Our RD analyses showed a similar pattern as Experiment 1. Consistent with Experiment 1, there was a main effect of WT whereby RD is shorter for the lexically read EXCs (*EXC RD* < *REG RD*), and approaches significance for INST (*name words RD* < *name all RD*), further supporting our hypotheses about shorter RDs being associated with lexically read items. Larger WF effects were again associated with the longer RD cells (i.e., REGs), despite the fact that the opposite pattern was demonstrated for onset RT. Our analysis of the SND effect on RD showed no significant effects.

The consistency in the results between Experiment 1 and 2 is reassuring, given that a voice-key was used to collect onset RT in Experiment 2, whereas hand-marking of onset RT was used in Experiment 1 (see **Figures 6** and **8**; cf. Rastle and Davis, 2002). Given that Experiment 1 was conducted in a MRI, we could not use a voice-key, but the gradient noise from the MRI scanner served as an effective auditory cue for identifying stimulus onset on the recording in that it was synchronized to appear coincidently with the last image acquisition prior to the gap for responding. In Experiment 2, we used a voice-key for detecting onset RT as we did not include an auditory cue for stimulus onset.

ANALYSIS OF COMBINED EXPERIMENTS 1 AND 2 DATA

WORD NAMING REACTION TIME

The data from the two experiments were combined, and the results were analyzed for all participants together ($n = 59$). The median naming onset RTs are presented in **Figure 11**. There were significant main effects of INST, [$F_{(1, 58)} = 31.7$, $MSe = 3009.1$, $p < 0.001$], and WT, [$F_{(1, 58)} = 8.6$, $MSe = 528.2$, $p = 0.005$], and no significant interaction, [$F_{(1, 58)} = 0.03$, $MSe = 303.5$, $p = 0.85$].

NON-WORD NAMING REACTION TIME

The NW condition in the *name all* INST condition yielded a median onset RT of 729.4 ms (Loftus and Masson, 1994, repeated-measure 95% CI = ±12.1).

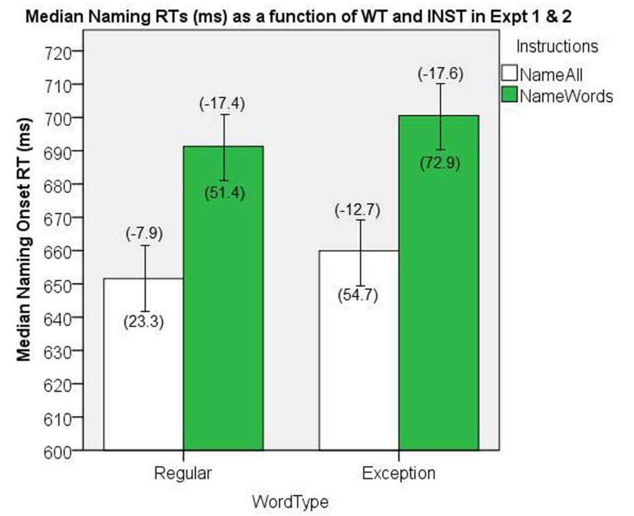


FIGURE 11 | Median Naming RTs (in ms) as a function of WT and INST in the combined analysis of Experiments 1 and 2. The 95% C.I.s are presented as error bars using Loftus and Masson's (1994) method. Co-efficients relating WF to RT (ms/$\log_{10}$HALWF) are presented in parentheses above each error bar, and co-efficients relating SND to RT (ms/unit inverseNcount) are presented in parentheses below each error bar.

ACCURACY

Given that there was no variance in Experiment 1 accuracy data, we did not perform a combined analysis of the Experiments 1 and 2 data.

WORD FREQUENCY EFFECTS ON REACTION TIME

We conducted analyses of WF effects on RT in the same way as our analyses in Experiments 1 and 2. **Figure 11** shows the mean co-efficients above the median RTs. There was a significant main effect of INST on the size of WF effect, [$F_{(1,\ 58)} = 23.4$, $MSe = 129.2$, $p < 0.001$], which represents an INST by WF over-additive interaction on naming RT, whereby the WF effects are greater under *name words* INST. The main effect of WT on the size of the WF effect approached significance, [$F_{(1,\ 58)} = 3.18$, $MSe = 112.6$, $p = 0.08$], which suggests a WT by WF over-additive interaction on naming RT averaging over both levels of the INST manipulation, whereby the WF effects are greater for EXCs than for REGs. More importantly, the analysis of this interaction under the normal *name all* INST condition yielded a significant effect, [$t_{(58)} = -2.49$, $SEM = 1.91$, $p = 0.02$][2]

SEMANTIC NEIGHBORHOOD DENSITY EFFECTS ON REACTION TIME

We conducted analyses of SND effects on RT in the same way as our analyses in Experiments 1 and 2. **Figure 11** shows the mean co-efficients below the median RTs. There was a significant main effect of INST on the size of the SND effect, [$F_{(1,\ 58)} = 10.08$, $MSe = 3134.9$, $p = 0.002$], which represents an INST by SND interaction on naming RT, whereby the SND co-efficients are greater in the *name words* INST condition than in the *name all* condition. There was also a significant main effect of WT on the size of the SND effect, [$F_{(1,\ 58)} = 9.70$, $MSe = 4264.6$, $p = 0.003$], which represents a SND by WT interaction on naming RT, whereby the SND co-efficients are greater for EXCs than for REGs.

WORD FREQUENCY AND SEMANTIC NEIGHBORHOOD DENSITY JOINT EFFECTS ON REACTION TIME

A GLM regression was conducted on each participant's correct onset RTs, with RT as the dependent variable, and SND and WF as continuous independent variables, separately for each combination of INST and WT. Given that this constitutes a multiple regression, we note that, here and in later multiple regression analyses, the issue of multicolinearity was handled by using a tolerance threshold set at 0.0001, and there were no situations whereby this threshold was exceeded. Multivariate outliers were assessed using Malhalanobis distance, and there were no multivariate outliers exceeding the threshold of [$\chi^2_{(2)} = 13.816$, $p < 0.001$]. The resulting WF × SND co-efficients for each INST and WT set were then tested against zero using one-sample t-tests. The WF × SND co-efficients were significantly different from zero for REGs in the *name all* INST condition, [$t_{(58)} = -2.40$, $SEM = 7.96$, $p = 0.02$], and for REGs in the *name words* INST condition, [$t_{(58)} = -2.11$, $SEM = 8.46$, $p = 0.04$]. The WF × SND co-efficients for EXCs in the *name words* INST condition approached significance, [$t_{(58)} = -1.77$, $SEM = 8.65$, $p = 0.08$], and the co-efficients for EXCs in the *name all* INST condition were not significant, [$t_{(58)} = -1.21$, $SEM = 5.76$, $p = 0.23$].

WORD NAMING RESPONSE DURATION

The median naming RDs are presented in **Figure 12**. There was a significant main effect of WT, [$F_{(1,\ 58)} = 257.6$, $MSe = 167.7$, $p < 0.001$], and of INST, [$F_{(1,\ 58)} = 4.39$, $MSe = 1023.9$, $p = 0.04$], and no significant interaction, [$F_{(1,\ 58)} = 1.39$, $MSe = 156.6$, $p = 0.24$].

NON-WORD NAMING RESPONSE DURATION

The NW condition in the *name all* INST condition yielded a median RD of 570.4 ms (Loftus and Masson, 1994, repeated-measure 95% CI = ±6.72).

WORD FREQUENCY EFFECTS ON RESPONSE DURATION

We conducted analyses of WF effects on RD in the same way as our analyses in Experiments 1 and 2. The main effect of INST on the size of WF effect approached significance, [$F_{(1,\ 58)} = 3.55$,

[2]We are concentrating on Two-Way interactions in this research given the focus on the Additive Factors Method. A reviewer had pointed out an interesting potential three-way interaction whereby the WF × WT interaction was only significant in the *name all* condition, but not the *name words* condition. Given that the WF effects are consistently negative for all of the conditions, any such Three-Way interactions would be ordinal (i.e., not a cross-over interaction), which are notoriously difficult to detect (i.e., all effects are in the same general direction). Nonetheless, we did examine tests for Three-Way interactions under the conditions in our study that had the most power to detect such interactions. We tested the WF × WT × INST interaction on RT, and it yielded the following result, [$F(1,\ 38) = 1.67$, $MSe = 160.40$, $p = 0.204$]. We also examined the corresponding interaction in the item analyses, and also found a non-significant result, [$F(1,\ 248) = 0.43$, $MSe = 2057.89$, $p = 0.51$].

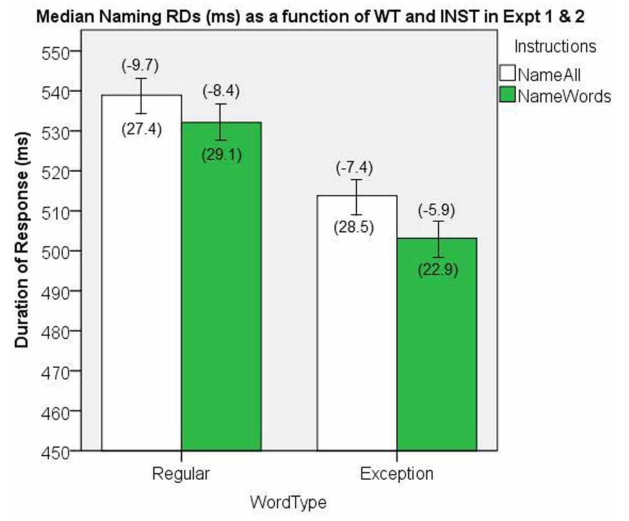


FIGURE 12 | Median Naming RDs (in ms) as a function of WT and INST in the combined analysis of Experiments 1 and 2. The 95% C.I.s are presented as error bars using Loftus and Masson's (1994) method. Co-efficients relating WF to RT (ms/$\log_{10}$HALWF) are presented in parentheses above each error bar, and co-efficients relating SND to RT (ms/unit inverseNcount) are presented in parentheses below each error bar.

MSe = 34.7, $p = 0.065$], which suggests a INST by WF interaction on naming RD, whereby the WF effects are greater for the *name all* INST condition. The main effect of WT on the size of WF effect was significant, [$F_{(1,\ 58)} = 9.77$, *MSe* = 35.0, $p = 0.003$], which represents a WT by WF interaction on naming RD, whereby the WF effects are greater for REGs. **Figure 12** shows the mean co-efficients above the median RDs.

SEMANTIC NEIGHBORHOOD DENSITY EFFECTS ON RESPONSE DURATION

In order to evaluate the effects of SND as a continuous variable, GLM regressions were conducted on each participant's correct RDs, as in Experiments 1 and 2. **Figure 12** shows the mean co-efficients below the median RDs. There was no significant main effect of INST on the size of the SND effect, [$F_{(1,\ 58)} = 0.145$, *MSe* = 1429.3, $p = 0.70$], nor was there a significant main effect of WT on the size of the SND effect, [$F_{(1,\ 58)} = 0.509$, *MSe* = 742.5, $p = 0.48$].

WORD FREQUENCY AND SEMANTIC NEIGHBORHOOD DENSITY JOINT EFFECTS ON RESPONSE DURATION

A GLM regression was conducted on each participant's correct onset RDs, with RD as the dependent variable, and SND and WF as continuous independent variables, separately for each combination of INST and WT. The resulting WF × SND co-efficients for each INST and WT set were then tested against zero using one-sample *t*-tests. The WF × SND co-efficients were significantly different from zero in the *name all* INST condition, for both EXCs, [$t_{(58)} = -6.80$, *SEM* = 2.74, $p < 0.001$], and REGs, [$t_{(58)} = -3.12$, *SEM* = 8.28, $p = 0.003$]. The WF × SND co-efficients for EXCs in the *name words* INST condition approached significance, [$t_{(58)} = -1.94$, *SEM* = 3.82, $p = 0.057$], and the co-efficients for REGs in the *name words* INST condition were not significant, [$t_{(58)} = -1.13$, *SEM* = 4.44, $p = 0.26$].

ITEM ANALYSES FOR REACTION TIME

An item analysis was performed on the combined data from Experiments 1 and 2. Median onset RTs for both *name all* and *name words* INST were treated as repeated measure dependent variables, and regressed on WT and INST using a repeated measure GLM. Given that our measure of WF was moderately correlated with our measure of SND, $r = -0.674$ when we use the inverse Ncount measure, as we have in these analyses ($r = 0.861$ when the non-inverse Ncount measure is used), we chose to analyze the effects of WF and semantic density in separate regression models, as well as together in a subsequent model so that we could assess all of the Two-Way joint effects.

In the regression model that included WF but not SND, there was a significant main effect of INST, [$F_{(1,\ 248)} = 113.38$, *MSe* = 2057.9, $p < 0.001$], and a main effect of WT, [$F_{(1,\ 248)} = 4.53$, *MSe* = 3896.8, $p = 0.03$]. There was also a significant main effect of WF, [$F_{(1,\ 248)} = 104.33$, *MSe* = 3896.8, $p < 0.001$]. There was no significant interaction between INST and WT, [$F_{(1,\ 248)} = 0.289$, *MSe* = 2057.9, $p = 0.59$]. There was a significant INST by WF interaction, [$F_{(1,\ 248)} = 50.05$, *MSe* = 2057.9, $p < 0.001$]. The WT by WF interaction approached significance by a one-tailed test, [$F_{(1,\ 248)} = 2.53$, *MSe* = 3896.8, $p = 0.057$].

In the regression that included SND but not WF, there was a significant main effect of INST, [$F_{(1,\ 248)} = 118.50$, *MSe* = 2176.0, $p < 0.001$]. There was no significant main effect of WT, [$F_{(1,\ 248)} = 0.27$, *MSe* = 4562.7, $p = 0.60$]. There was a significant main effect of SND, [$F_{(1,\ 248)} = 51.45$, *MSe* = 4562.7, $p < 0.001$]. There was no significant interaction between INST and WT, [$F_{(1,\ 248)} = 1.24$, *MSe* = 2176.0, $p = 0.26$]. There was a significant interaction between INST and SND, [$F_{(1,\ 248)} = 34.0$, *MSe* = 2176.0, $p < 0.001$]. There was a significant WT by SND interaction, [$F_{(1,\ 248)} = 4.13$, *MSe* = 4562.7, $p = 0.043$].

In the regression that included both WF and SND, there was a significant main effect of INST, [$F_{(1,\ 245)} = 24.89$, *MSe* = 2034.2, $p < 0.001$]. There was no main effect of WT, [$F_{(1,\ 245)} = 0.01$, *MSe* = 3757.9, $p = 0.94$]. There was a significant main effect of SND, [$F_{(1,\ 245)} = 9.38$, *MSe* = 3757.9, $p = 0.002$], and a significant main effect of WF, [$F_{(1,\ 245)} = 27.83$, *MSe* = 3757.9, $p < 0.001$]. There was no significant interaction between INST and WT, [$F_{(1,\ 245)} = 0.01$, *MSe* = 2034.2, $p = 0.93$]. There was a significant interaction between INST and SND, [$F_{(1,\ 245)} = 4.23$, *MSe* = 2034.2, $p = 0.041$]. There was a significant interaction between INST and WF, [$F_{(1,\ 245)} = 10.01$, *MSe* = 2034.2, $p = 0.002$]. The WT by SND interaction approached significance, [$F_{(1,\ 245)} = 3.43$, *MSe* = 3757.9, $p = 0.065$] (which, given the significant interaction in our earlier analyses, could be assessed by a one-tailed test with $p = 0.0325$). There was no significant WT by WF interaction, [$F_{(1,\ 245)} = 0.01$, *MSe* = 3757.9, $p = 0.93$]. There was a significant interaction between SND and WF, [$F_{(1,\ 245)} = 8.24$, *MSe* = 3757.9, $p = 0.004$].

ITEM ANALYSES FOR RESPONSE DURATION

An item analysis was performed on the combined data from Experiments 1 and 2. Median RDs for both *name all* and *name words* INST were treated as repeated measure dependent variables, and regressed on WT and INST using a repeated measures GLM.

In the regression model that included WF but not SND, there was a significant main effect of WF, [$F_{(1, 248)} = 20.9$, $MSe = 5249.7$, $p < 0.001$]. There were no other significant main effects or interactions.

In the regression model that included SND but not WF, there was a significant main effect of INST, [$F_{(1, 248)} = 9.21$, $MSe = 1043.1$, $p = 0.003$], a significant main effect of WT, [$F_{(1, 248)} = 7.21$, $MSe = 5518.0$, $p = 0.008$], and a significant main effect of SND, [$F_{(1, 248)} = 7.83$, $MSe = 5518.0$, $p = 0.006$]. There were no significant Two-Way interactions.

In the regression that included both WF and SND, the only significant main effect was WF, [$F_{(1, 245)} = 6.75$, $MSe = 5230.4$, $p = 0.01$], and the main effect of SND approached significance, [$F_{(1, 245)} = 3.10$, $MSe = 5230.4$, $p = 0.08$]. The only Two-Way interaction that approached significance was between SND and WF, [$F_{(1, 245)} = 3.56$, $MSe = 5230.4$, $p = 0.06$]. There were no other significant main effects or interactions.

GENERAL DISCUSSION

As described in the introduction, our first set of hypotheses involved the joint effects of INST, WF, SND, and WT on onset RT. Taken together, the by-participants analyses, the by-item-by-participant regression analyses, and the by-item analyses, support our hypotheses such that: (1) *INST* × *WF*—the INST × WF over-additive interaction was significant, supporting the notion that these two variables are affecting the orthographic lexical system; (2) *INST* + *WT*—given that INST should be having its effect in the early stages of word processing, whereas WT has previously been shown to have its effect at a later stage of processing, this additive pattern was also consistent with our previous research (Cummine et al., 2010, 2012; Borowsky et al., 2012), and supports the notion that INST are affecting an orthographic lexical system that is temporally separable from the phonological output system that is affected by WT; (3) *INST* × *SND*—the INST × SND over-additive interaction was significant, supporting a common locus of the orthographic lexical system for their effects; (4) *WF* × *SND*—the WF × SND over-additive interaction supports the notion that these two variables are affecting the semantic system and the connections to other word-level systems; (5) *SND* × *WT*—the SND × WT over-additive interaction was significant, supporting a common locus of the phonological output system for their effects; and (6) *WF* × *WT*—the WF × WT over-additive interaction under the normal *name all* INST was significant in Experiment 2 and the combined analyses, supporting the notion that these variables affect the phonological output system.

Our second set of hypotheses involved the RDs of vocalizations: *EXC RD* < *REG RD* < *NW RD*—given that EXCs must be processed as whole-words and read lexically in order to be pronounced correctly, they produced the shortest RDs, despite the fact that EXC onset RTs are longer than REG onset RTs. Given that NWs must be processed through sublexical GPCs, they produced the longest RDs. Finally, given that REGs can be processed through either route, they elicited intermediate RDs relative to EXCs and NWs, despite having the fastest onset RTs. The results supported the prediction that *name words RD* < *name all RD* in that the more lexically a word is read, the shorter the RD. The SND effect remained in the same facilitative direction for onset RT and RD, which is consistent with Balota et al.'s (1989) finding with semantic priming. The WF effect also remained in the same facilitative direction for onset RT and RD, which is consistent with it having its effects at the same lexical/semantic level as SND.

RESPONSE DURATION

By developing a new measure of RD for reading aloud, we have an additional and more comprehensive measure of basic reading processes. Given that basic reading processes are still ongoing after the initiation of a vocal response, measures of onset RT may only reflect early aspects of processing (e.g., in terms of only *partially* reflecting lexical access, or the resolution of conflicting phonological codes). Our results provide evidence that systems that are influenced by such variables as INST, WF, SND, and WT are still affecting the duration of the reading response, even after these variables have already influenced onset RTs. In addition, our results support the notion that the more lexically a word is read (e.g., EXCs, or *name words* INST), the shorter the RD, in the face of longer onset RTs for such conditions (see **Figures 13** and **14**). To our knowledge, this is the first demonstration of dissociation between RT and RD, as a

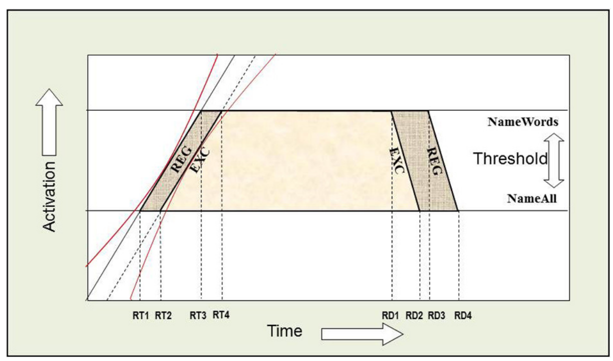


FIGURE 13 | Additive joint effects on RT and RD between INST and WT. An Additive Factors Method (AFM) interpretation of the additive effect on RT is that INST and WT are affecting separable systems in time. If INST are assumed to affect the threshold for activation (i.e., the amount of time it takes to verify that a letter string spells a word), and WT is assumed to shift the rate of activation over time (i.e., the time it takes to choose among the competing phonological codes for EXCs), or vice versa, then the points in time when each rate crosses a threshold correspond to the average onset RTs (the left side of each trapezoid). The red confidence intervals around the slopes allow for trial to trial variation as described by Masson and Kliegl (2012), and thus we note that such variation is not problematic for applying the AFM, and can be easily accommodated by cascaded stages of processing. The effects of INST and WT on RD represent a dissociation when compared to RT. The right side of each trapezoid represents the RD and illustrates the dissociation.

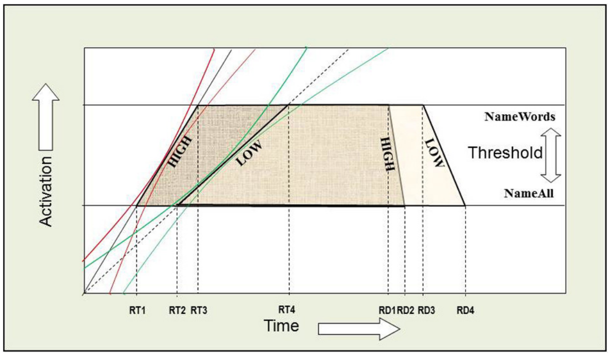


FIGURE 14 | Over-additive interaction on RT between INST and WF. An AFM interpretation of this interaction on RT is that both INST and WF are affecting a common system in time. If INST are assumed to affect the threshold for activation and WF is assumed to affect the rate of activation over time, or vice versa, then the points in time when each rate crosses a threshold correspond to the average onset RTs (the left side of each trapezoid). WF and WT interact on RT in a similar fashion. The red and green confidence intervals around the slopes allow for trial to trial variation. The effects of WF on RD remain negative when compared to RT. We note that SND effects can be accounted for in a similar manner. The effects of INST on RD represent the dissociation when compared to RT. The right side of each trapezoid represents the RD and illustrates this dissociation.

function of the degree of lexical-based reading. This dissociation is particularly powerful given that it has been demonstrated by both a within-item (i.e., INST) and between-item (i.e., WT) manipulation. Although Balota et al. (1989) showed that semantic priming had an effect on both RT and RD in their study, whereby both were shorter for the related condition, WT and INST in the present study are clearly showing a dissociation between onset RT and RD. Interestingly, neither SND nor WF effects reversed between RT and RD (i.e., inverse Ncount co-efficients remained positive, while WF co-efficients remained negative in all conditions), and thus seem to be behaving in a manner consistent with semantic priming effects (**Figure 14**).

Perhaps SND and WF effects (similar to Balota et al.'s, 1989 finding with semantic priming) can be thought of as consistently reflecting the core lexical/semantic aspects of processing, in that they both facilitate the speed of lexical/semantic access and thus affect onset RT and RD in the same way. Instruction effects might best be thought of as reflecting a front-end gating manipulation, whereby INST to *name words* serves to increase reliance on the orthographic lexical route, relative to INST to *name all*. RT is higher for *name words* INST given the time required to verify the word's lexical status, whereas RD is shorter in this condition given that once the word's lexical status has been verified, the phonological code is also lexical-based and thus more rapidly produced. However, it is clear that EXCs are showing different RDs for the two INST conditions, suggesting that there is "room to move" for EXCs to have even shorter RDs under the *name words* INST condition compared to the *name all* INST condition. Perhaps under *name all* INST, EXC RDs are produced with some hesitation due to the greater overall reliance on sublexical GPCs under *name all* INST (e.g., for naming the NWs, and perhaps REGs some of the time). Word-type effects might best be thought to reflect a back-end convergence effect, whereby both routes produce converging phonological codes for REGs, relative to EXCs whereby both routes produce conflicting phonological codes. RT is shorter for REGs given an early assessment of the phonological codes for the word's onset, allowing a participant to quickly initiate their response, whereas RD is longer given the slower sublexical contribution to completing the entire word's pronunciation.

The present RD results also bear on the question of the degree to which reading processes are still occurring post-vocalization onset. Given the significant effects of INST, WF, SND, and WT on RD, there is clearly a substantial amount of processing occurring post-vocalization onset. These post-vocalization onset effects clearly support the utility of an RD measure for investigating reading processes.

ADDITIVE FACTORS METHOD AND MODELS OF READING

The AFM allows for the investigation of whether two or more variables are affecting common or separable systems in time, whereby two variables that interact over-additively on RT are considered to affect a common system, whereas two variables that are additive on RT are considered to affect separable systems. We selected four lexical/semantic variables known to affect basic reading processes, and examined their joint effects so as to delineate the sequence of systems involved in reading aloud. We manipulated: INST as a variable that would serve to gate processing toward the lexical route when participants were to *name words* only; WF as a variable whose effects reflect lexical/semantic connections; SND as a semantic variable whose effects reflect associations among semantic neighbors and their lexical/semantic connections; and WT as a variable whose effects reflect the convergence of the sublexical and lexical routes, in that REGs can be pronounced correctly through either route, whereas EXCs create conflicting phonological codes through the two routes. The INST × WF and INST × SND over-additive interactions support the notion that these variables are affecting a common and relatively early system in time, namely the orthographic lexical system (see **Figure 1**). The WF × WT and SND × WT over-additive interactions support the notion that these variables affect a common and relatively later system in time, namely the phonological output system. The INST + WT additive joint effects support the notion that they are affecting the orthographic lexical and phonological output systems, respectively. Taken together, these joint effects clearly support a model where the orthographic lexical system and phonological output system are cascaded in time (**Figure 1**), and not operating in parallel (cf., Seidenberg et al., 1996; Plaut and Booth, 2000).

The issue of the naming task being less sensitive to semantic effects, as described in the Introduction (Borowsky and Masson, 1996; Yap et al., 2012), is also addressed with the present results. By instructing readers to pronounce words only after lexical verification (i.e., the *name words* INST condition), and thus requiring them to read lexically, there was a larger SND effect than when they were instructed to name without encouraging reliance on the lexical route (i.e., the *name all* INST condition). As such, the INST × SND over-additive interaction supports the notion that semantic influences on naming behavior can occur under conditions that encourage lexical access. The SND × WT over-additive interaction also supports this notion, whereby EXCs, which must be processed via the orthographic lexical route, showed a larger SND effect than REGs[3].

VENTRAL-LEXICAL, DORSAL-SUBLEXICAL, MODEL OF BASIC READING PROCESSES

Our preferred cognitive architecture for basic reading processes, which is based on the Dual-Route Cascade model (Coltheart et al., 2001), assumes that processing operates on two routes: a sublexical GPC route, which allows less familiar letter strings (including NWs and novel words) to be "sounded out," and a lexical route, which allows familiar words to be read as whole-words (see **Figure 1**). These routes have been mapped onto the dorsal and ventral visual processing streams, respectively (Herbster et al., 1997; Jobard et al., 2003; Price and Devlin, 2003; Indefrey and Levelt, 2004; Joubert et al., 2004; Borowsky et al., 2006, 2007, 2012; Hickok and Poeppel, 2007; Cohen et al., 2008; Cummine et al., 2010, 2012). This model also assumes that processing is cascaded among the subsystems.

The ventral-lexical route is relied on for reading familiar REGs and EXCs. The dorsal-sublexical route is relied on for reading NWs, novel words, and less familiar REGs. The convergence of these two routes can be facilitative in the case of reading REGs (where the phonological codes would be the same from both routes), or conflicting in the case of reading EXCs (where the phonological codes would be different from both routes), which has been described earlier in the context of the interaction between WF and WT (see also Cummine et al., 2010). Given that WF and SND have their effects in the lexical/semantic systems, including the orthographic lexical system, their influences are early enough to interact with the effects of INST, which can gate processing through the lexical route under the *name words* condition. In order to allow for novel words that lack any lexical representation to be read aloud, there is also a pathway from GPC to phonological output (see also Borowsky et al., 2002).

It is worth noting that the ventral-lexical, dorsal-sublexical, multiple stage model and the effects of INST and WT that we are describing here are also consistent with other findings in the literature that have underscored the necessity of multiple stages and attentional control in basic reading models. For example, Reynolds and Besner (2005) showed that participants took longer to name both words and NWs when the item on the preceding trial was from the other lexical category, relative to when the preceding item was from the same lexical category, which is similar to our account for why EXCs show different RDs under the two INST conditions (see also Reynolds and Besner, 2006, for a multiple stage account of attention and reading processes). Furthermore, Reynolds and Besner (2011) have also showed changes in the WF effects as a function of list context when reading pseudohomophones aloud (see also Borowsky et al., 2002).

SINGLE-MECHANISM MODELS

Single-mechanism parallel distributed processing (PDP) models (e.g., Plaut and Booth, 2000) are challenged by the current results. Such models have been developed to account for the basic effect of WT (REG, EXC), and the WF by WT interaction, on RT by a "division-of-labor" between an Orthography–Phonology (O–P) pathway and an Orthography–Semantics–Phonology (O–S–P) pathway (e.g., Harm and Seidenberg, 2004, although "pathway" may be misleading given that these models subscribe to parallel processing across the entire network). Larger WF effects for EXCs are thought to occur due to the additional WF-sensitive connections involved in the O–S–P pathway, compared to the O–P pathway that REG reading is thought to rely on. There is no distinct orthographic lexicon in these models, unlike the dual-stream models, and so INST to read by first checking the orthographic lexicon (name words, based on spelling) raises a challenge in and of itself. Waiting (to any degree) for the O units to settle on the word's pattern of activation and using that information to gate processing in the S and P units might be a solution, but such "stage-like" or "cascaded" processing is counter to the *parallel* definition of these models (Plaut and Booth, 2000; and see the debate by Borowsky and Besner, 2006; Plaut and Booth, 2006, and Besner and Borowsky, 2006, for additional discussion of these issues, and see Ziegler et al., 2009, for a more recent hybrid computational model that has implemented thresholds or "stages of processing" in order to account for some additive effects). Perhaps most challenging is the presence of additive effects on RT (i.e., INST and WT) in the same range of RTs that also show over-additive interactions. Although a sigmoid activation function within a single-mechanism PDP model has been explored as a means to account for both additivity and over-additive interactions (Plaut and Booth, 2000), this approach is problematic as additive effects can only arise equidistant from the center of the sigmoid input–output function, yet additive effects occur regularly within the very same range of RTs as over-additive effects, as demonstrated in the research reported here, and elsewhere (see Borowsky and Besner, 2006 and Cummine et al., 2012 for a review).

Additive effects in the same range of RTs as over-additive effects are still best accounted for by the AFM. Additive effects of two variables are easily accounted for by implementing the effects of the two variables at two different time points in processing (i.e., two systems with stage-like processing). These systems may be in cascade (e.g., McClelland and Rumelhart, 1981; Borowsky and Besner, 1993; Coltheart et al., 2001)—all that is necessary is at least some delay between the initiation of activation in one system compared to the other. Such a delay is parsimonious with the known behavior of real neural networks, and thus it can also be

[3] In response to a reviewer's query about SND effects under the *name all* INST condition, we conducted one-sample *t*-tests on the co-efficients that relate SND to RT and to RD. All SND co-efficients were found to be significantly different from zero (all $t_{(58)} > 4.25$, $p < 0.001$). We also examined WF effects in the same way, and found that all WF co-efficients were significantly different from zero (all $t_{(58)} > |-7.06|$, $p < 0.001$). These results support the previously described idea that both SND and WF effects consistently reflect the core lexical/semantic aspects of processing, in that they both facilitate the speed of lexical/semantic access and thus affect onset RT and RD in the same way.

argued to be a necessary characteristic in all neurobiological models. Over-additive effects of two variables can be accounted for by implementing the effects of the two variables within the same system of processing (e.g., by affecting its activation rate, threshold, or baseline activation level, see Borowsky and Besner, 1993, 2006, for discussion about how these parameters can be modeled to account for additivity and over-additivity). Dual-stream models of reading can readily handle the over-additive interactions as long as cascaded processing (i.e., some degree of delay of activation in systems down-stream) is assumed.

ADVANTAGES OF THE PRESENT METHOD

Despite the amount of time that goes into hand-marking each vocalization, the benefits from this approach far outweigh the costs (see also Rastle and Davis, 2002). Not only does hand-marking provide a new set of empirical data (RD) for testing models of reading processes, it also provides the following advantages. The traditional definition of a spoiled trial in a naming experiment includes a substantial proportion of trials when the voice-key failed to trigger—in the present experiments, the proportion of spoiled trials is quite low, given that replaying the audio is an important part of zeroing-in on the onset and offset, which is not done when one relies on a voice-key. Recording and hand-marking of vocal responses also allows for the collection of overt naming behavioral data in the MRI environment. Experiment 1 was conducted in the context of a fMRI study, and by simply recording through the intercom and synchronizing stimulus onset with an image acquisition, we were able to clearly detect onset and durations of vocal responses. By also using sparse sampling (i.e., a gap in image acquisition), the participants' vocal response was made in a relatively noise-free time period, which was also helpful. As such, recording and hand-marking of vocal responses will be of great benefit to researchers who do fMRI experiments involving vocalization responses. We note that there has been some computer software developed to analyze for onset and duration (e.g., Kello and Kawamoto, 1998), but such an approach would not be as effective as hand-marking and replaying vocal responses with respect to detecting spoiled responses and individual differences in intensity of vocalization, especially in noisy environments such as an MRI.

CONCLUSION

Pursuing an understanding of the meanings of things in our world is a central feature of the human condition. We have an insatiable curiosity to understand how to interact with objects in our environment, how to interpret symbols and actions, as well as the meanings of words. Given that semantic knowledge is core to understanding words, our present research explored the interactions between semantic and lexical variables in order to inform the development of a model of basic reading processes. The joint effects of INST, WF, SND, and WT on naming RT support a cascaded, dual-route, ventral-lexical/dorsal-sublexical model. Our naming RD results provide evidence that basic reading processes, and their joint effects, are occurring even after the initiation of a vocal response, and support the notion that the more lexically a word is read, the shorter the RD. Given the joint effects on RT, the dissociating effects of INST and WT on RT versus RD, and consistent effects of WF and SND on RT and RD, models of basic reading processes now have new challenges to accommodate these effects. An important question for future research is the degree to which RD effects are due to phonological-lexical vs. orthographic-lexical processing, which our lab is beginning to explore.

ACKNOWLEDGMENTS

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada in the form of a research assistantship to Layla Gould, and grants to Ron Borowsky and Jacqueline Cummine. Ron Borowsky is currently funded by NSERC Discovery Grant 183968-2008-2013. The lead author (Layla Gould) and senior author (Ron Borowsky) contributed equally to this work. Correspondence can be sent to layla.gould@usask.ca, or ron.borowsky@usask.ca.

REFERENCES

Abrams, R. A., and Balota, D. A. (1991). Mental chronometry: beyond reaction time. *Psychol. Sci.* 2, 153–157.

Balota, D. A., Boland, J., and Shields, L. (1989). Priming in pronunciation: beyond pattern recognition and onset latency. *J. Mem. Lang.* 28, 14–36.

Balota, D. A., and Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of the neglected decision stage. *J. Exp. Psychol. Hum. Percept. Perform.* 10, 340–357.

Balota, D. A., and Chumbley, J. I. (1985). The locus of word frequency effects in the pronunciation task: lexical access and/or production? *J. Mem. Lang.* 24, 89–106.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S., Spieler, D. H., and Yap, M. J. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. Gen.* 133, 283–316.

Balota, D. A., Yap, M. J., and Cortese, M. J. (2006). "Visual word recognition: the journey from features to meaning (a travel update)," in *Handbook of Psycholinguistics, 2nd Edn*, eds M. Traxler and M. A. Gernsbacher (Amsterdam, NY: Academic Press), 285–375.

Balota, D. A., Yap, M., Cortese, M., Hutchison, K., Kessler, B., Loftis, B., et al. (2007). The english lexicon project. *Behav. Res. Methods* 39, 445–459.

Bangert, A. S., Abrams, R. A., and Balota, D. A. (2012). Reaching for words and nonwords: interactive effects of word frequency and stimulus quality on the characteristics of reaching movements. *Psychon. Bull. Rev.* 19, 513–520.

Besner, D., and Borowsky, R. (2006). Plaut and Booth's new simulations: what have we learned? (Postscript to Borowsky and Besner, 2006). *Psychol. Rev.* 113, 194–195.

Boder, D. P. (1933). Some new electronic devices for the psychological laboratory. *Am. J. Psychol.* 45, 145–147.

Boersma, P., and Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.30, retrieved 6 October 2012 from http://www.praat.org/

Borowsky, R., and Besner, D. (1993). Visual word recognition: a multistage activation model. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 813–840.

Borowsky, R., and Besner, D. (2006). Parallel distributed processing and lexical-semantic effects in visual word recognition: are a few stages necessary? *Psychol. Rev.* 113, 181–195.

Borowsky, R., Cummine, J., Owen, W. J., Friesen, C. K., Shih, F., and Sarty, G. (2006). fMRI of ventral and dorsal processing streams in basic reading processes: insular sensitivity to phonology. *Brain Topogr.* 18, 233–239.

Borowsky, R., Esopenko, C., Cummine, J., and Sarty, G. (2007). Neural representations of visual words and objects: a functional MRI study on the modularity of reading and object processing. *Brain Topogr.* 20, 89–96.

Borowsky, R., Esopenko, C., Gould, L., Kuhlmann, N., Sarty, G., and Cummine, J. (2012). Localization of function for noun and verb processing: converging evidence for shared processing from fMRI activation and reaction time. *Lang. Cogn. Process.* doi: 10.1080/01690965.2012.665466. [Epub ahead of print].

Borowsky, R., Loehr, J., Friesen, C. K., Kraushaar, G., Kingstone, A., and Sarty, G. (2005a). Modularity and intersection of "what", "where", and "how" processing of visual stimuli: a new method of fMRI localization. *Brain Topogr.* 18, 67–75.

Borowsky, R., Owen, W., Wile, T., Friesen, C., Martin, J., and Sarty, G. (2005b). Neuroimaging of language processes: fMRI of silent and overt lexical processing and the promise of multiple process imaging in single brain studies. *Can. Assoc. Radiol. J.* 56, 204–213.

Borowsky, R., and Masson, M. E. J. (1996). Semantic ambiguity effects in word-identification. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 63–85.

Borowsky, R., Owen, W., and Masson, M. (2002). Diagnostics of phonological lexical processing: pseudohomophone naming advantages, disadvantages, and base-word frequency effects. *Mem. Cognit.* 30, 969–987.

Chumbley, J. I., and Balota, D. (1984). A word's meaning affects the decision in lexical decision. *Mem. Cognit.* 12, 590–606.

Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., and Montavont, A. (2008). Reading normal and degraded words: contribution of the dorsal and ventral visual pathways. *Neuroimage* 40, 353–356.

Coltheart, M. (2006). Dual route and connectionist models of reading: an overview. *Lond. Rev. Educ.* 4, 4–17. ∗ Special issue- The psychology of reading.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256.

Cummine, J., Sarty, G., and Borowsky, R. (2010). Localizating the Frequency x Regularity word reading interaction in the cerebral cortex. *Neuropsychologia* 48, 2147–2157.

Cummine, J., Gould, L., Zhou, C., Hrybouski, S., Siddiqi, Z., Chouinard, B., et al. (2012). Manipulating instructions strategically affects reliance on the ventral-lexical reading stream: converging evidence from neuroimaging and reaction time. *Brain Lang.* doi: 10.1016/j.bandl.2012.04.009. [Epub ahead of print].

Donders, F. C. (1969). "On the speed of psychological processes. (W. Koster, Trans.)," in *Attention and Performance, II.* ed W. Koster (Amsterdam: North-Holland).

Dunlap, K. (1913). Apparatus for association timing. *Psychol. Rev.* 20, 244–246.

Harm, M. W., and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111, 662–720.

Herbster, A., Mintun, M., Nebes, R., and Becker, J. (1997). Regional cerebral blood flow during word and nonword reading. *Hum. Brain Mapp.* 5, 84–92.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.

Hino, Y., and Lupker, S. (2000). Effects of word frequency and spelling-to-sound regularity in naming with and without preceding lexical decision. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 166–183.

Hudson, P., and Bergman, M. (1985). Lexical knowledge in word recognition: word length and word frequency in naming and lexical decision tasks. *J. Mem. Lang.* 24, 46–58.

Indefrey, P., and Levelt, W. (2004). The spatial and temporal signatures of word production components. *Cognition* 92, 101–144.

Jobard, G., Crivello, F., and Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *Neuroimage* 20, 693–712.

Joubert, S., Beauregard, M., Walter, N., Bourgouin, P., Beaudoin, G., Leroux, J., et al. (2004). Neural correlates of lexical and sublexical processes in reading. *Brain Lang.* 89, 9–20.

Kawamoto, A., Kello, C., Higareda, I., and Vu, J. (1999). Parallel processing and initial phoneme criterion in naming words: evidence from frequency effects on onset and rime duration. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 362–861.

Kello, C. T., and Kawamoto, A. H. (1998). Runword: an IBM-PC software package for the collection and acoustic analysis of speeded naming responses. *Behav. Res. Meth. Instrum. Comput.* 30, 371–383.

Kinoshita, S., and Woollams, A. M. (2002). The masked onset priming effect in naming: computation of phonology or speech-planning? *Mem. Cognit.* 30, 237–245.

Loftus, G. F., and Masson, M. E. J. (1994). Using confidence intervals in within-subjects designs. *Psychon. Bull. Rev.* 1, 476–490.

Masson, M. E. J., and Kliegl, R. (2012). Modulation of additive and interactive effects in lexical decision by trial history. *J. Exp. Psychol. Learn. Mem. Cogn.* doi: 10.1037/a0029180. [Epub ahead of print].

McCann, R. S., and Besner, D. (1987). Reading pseudohomophones: implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 14–24.

McClelland, J. L. (1979). On the time-relations of mental processes: an examination of systems of processes in cascade. *Psychol. Rev.* 86, 287–330.

McClelland, J. L., and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: part 1. An account of basic findings. *Psychol. Rev.* 5, 375–407.

Monsell, S., Doyle, M. C., and Haggard, P. N. (1989). The effects of frequency upon visual word recognition: where are they? *J. Exp. Psychol. Gen.* 118, 43–71.

Monsell, S., Patterson, K., Graham, A., Hughes, C., and Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: strategic anticipation of lexical status. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 452–467.

Morton, J. (1979). "Facilitation in word recognition: experiments causing change in the logogen model," in *Processing of Visible Language,* eds P. A. Kolers, M. E. Wrolstad, and M. Bouma. (New York, NY: Plenum). 259–268.

Patterson, K., and Hodges, J. (1992). Deterioration of word meaning: implications for word reading. *Neuropsychologia* 30, 1025–1040.

Plaut, D. C., and Booth, J. R. (2000). Individual and developmental differences in semantic priming: empirical and computational support for a single-mechanism account of lexical processing. *Psychol. Rev.* 107, 786–823.

Plaut, D. C., and Booth, J. R. (2006). More modeling but still no stages: reply to Borowsky and Besner. *Psychol. Rev.* 113, 196–200.

Price, C., and Devlin, J. (2003). The myth of the visual word form area. *Neuroimage* 19, 473–481.

Rastle, K., and Coltheart, M. (1999). Serial and strategic effects in reading aloud. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 482–503.

Rastle, K., and Davis, M. (2002). On the complexities of measuring naming. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 307–314.

Rastle, K., Harrington, J., Coltheart, M., and Palethorpe, S. (2000). Reading aloud begins when the computation of phonology is complete. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1178–1191.

Reynolds, M., and Besner, D. (2005). Basic processes in reading: a critical review of pseudohomophone effects in reading aloud and a new computational account. *Psychon. Bull. Rev.* 12, 622–646.

Reynolds, M., and Besner, D. (2006). Reading aloud is not automatic: processing capacity is required to generate a phonological code from print. *J. Exp. Psychol. Hum. Percept. Perform* 32, 1303–1323.

Reynolds, M., and Besner, D. (2011). There goes the neighbourhood: contextual control over the breadth of lexical activation when reading aloud. *Q. J. Exp. Psychol. (Hove)* 64, 2405–2424.

Roberts, S., and Sternberg, S. (1993). The meaning of additive reaction-time effects: tests of three alternatives. *Atten. Perform.* 14, 611–653.

Shaoul, C., and Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behav. Res. Methods* 42, 393–413.

Seidenberg, M. S., Petersen, A., MacDonald, M. C., and Plaut, D. C. (1996). Pseudohomophone effects and models of word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 48–62.

Sternberg, S. (1969). The discovery of processing stages: extensions of Donders' method. *Acta Psychologica* 30, 276–315.

Stolz, J., and Neely, J. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 3, 596–611.

Yap, M. J., and Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 274–296.

Yap, M. J., and Balota, D. A. (2009). Visual word recognition of multisyllabic words. *J. Mem. Lang.* 60, 502–529.

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., and Huff, M. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:72. doi: 10.3389/fnhum.2012.00072

Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.

Zevin, J., and Balota, D. (2000). Priming and attentional control of lexical and sublexical pathways during naming. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 121–135.

Ziegler, J. C., Perry, C., and Zorzi, M. (2009). Additive and interactive effects of stimulus degradation: no challenge for CDP. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 306–311.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 July 2012; accepted: 28 September 2012; published online: 19 October 2012.

Citation: Gould L, Cummine J and Borowsky R (2012) The cognitive chronometric architecture of reading aloud: semantic and lexical effects on naming onset and duration. Front. Hum. Neurosci. ***6****:287. doi: 10.3389/fnhum.2012.00287*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 27 November 2012
doi: 10.3389/fnhum.2012.00315

The semantic richness of abstract concepts

Gabriel Recchia*[1] *and Michael N. Jones*[1,2]*

[1] *Department of Cognitive Science, Indiana University, Bloomington, IN, USA*
[2] *Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA*

Edited by:
Melvin Yap, National University of Singapore, Singapore

Reviewed by:
Lara Jones, Wayne State University, USA
Jen Coane, Colby College, USA

****Correspondence:***
Michael N. Jones, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA.
e-mail: jonesmn@indiana.edu

We contrasted the predictive power of three measures of semantic richness—number of features (NFs), contextual dispersion (CD), and a novel measure of number of semantic neighbors (NSN)—for a large set of concrete and abstract concepts on lexical decision and naming tasks. NSN (but not NF) facilitated processing for abstract concepts, while NF (but not NSN) facilitated processing for the most concrete concepts, consistent with claims that linguistic information is more relevant for abstract concepts in early processing. Additionally, converging evidence from two datasets suggests that when NSN and CD are controlled for, the features that most facilitate processing are those associated with a concept's physical characteristics and real-world contexts. These results suggest that rich linguistic contexts (many semantic neighbors) facilitate early activation of abstract concepts, whereas concrete concepts benefit more from rich physical contexts (many associated objects and locations).

Keywords: concreteness, lexical decision, semantic richness, feature norms, abstract concepts

The majority of experimental evidence guiding our knowledge of lexical semantic representation comes from research using concrete words. Concrete words are the dominant stimuli in the literatures on semantic priming, property generation, and single-item recognition (e.g., naming or lexical decision in isolation). In contrast, a larger proportion of the human lexicon may actually be composed of abstract words. For example, of the 500 most frequent words in the TASA corpus (Landauer and Dumais, 1997), 70% are classified as abstract according to the MRC Psycholinguistic Database (Coltheart, 1981). In studies involving large representative samples of English nouns, a substantial proportion (over 40%) are rated as abstract by human raters (Gilhooly and Logie, 1980) and classified as abstract entities in lexical databases such as WordNet (Miller, 1990). Furthermore, abstract words have been implicated as particularly worthy of study due to their status as network hubs (Sigman and Cecchi, 2002), the challenge they pose to grounded theories of cognition (Barsalou, 2008), and their higher relative frequency than concrete words (Audet and Burgess, 1999). In his review of research on grounded cognition, Barsalou (2008) notes "Because the scientific study of concepts has primarily focused so far on concrete concepts, we actually know remarkably little about abstract concepts, even from the perspective of traditional cognitive theories" (p. 634). The disconnect between the type of words we know the most about and the type of words that most inhabit the lexicon means that theoretical development may be over-emphasizing mechanisms and information sources for word representation that do not generalize to the full lexicon.

Abstract concepts are particularly important to study because various theories of semantic representation make different claims about the degree of semantic richness possessed by abstract concepts, particularly with respect to semantic features. Abstract concepts may be semantically impoverished, deriving their meaning primarily from their associations with other words (Paivio, 1986, 2010; Plaut and Shallice, 1993). Alternatively, they may be no less semantically rich, but be more grounded in introspective simulations (Barsalou et al., 2008) or aspects of meaning related to their social/communicative function (Borghi and Cimatti, 2009; Borghi et al., 2011).

If semantic features are important to abstract concept representations, one might expect to find feature-based effects similar to those observed for concrete words. For concrete concepts, being associated with many semantic features facilitates lexical processing (Pexman et al., 2002, 2003; Grondin et al., 2009). These so-called *number of feature* (NF) effects have established the importance of semantic richness in concrete word representation. Investigating whether NF effects are obtained for abstract words—and if so, for what types of features—can yield insight into their representations and the information sources used to learn those representations.

Pexman et al. (2008) found that in addition to NF, a concept's number of semantic neighbors (NSN) and contextual dispersion (CD) accounted for unique response time variance in a lexical decision task. However, their reliance upon the McRae et al. (2005) feature norms to calculate NF restricted their analysis to concrete words. In this paper, we use a novel online game modeled after McRae et al.'s task to gather feature generation data, and present results from data collected from 30 subjects/word for 550 words, including 177 abstract concepts. Extending the methods of Pexman et al. (2008) to this database and to alternative measures of NSN, NF, and CD, we evaluate whether NSN, NF, and CD each account for unique variance in lexical decision times (LDT) for abstract as well as concrete words. We also investigated the specific types of features that contribute to NF effects when NSN and CD are controlled for.

ARE ABSTRACT CONCEPTS RICH IN ANYTHING?

Several studies have investigated whether the processing and memory advantages often observed for concrete words are due to their allegedly richer featural representations (e.g., Saffran, 1980; Barry, 1984; Plaut and Shallice, 1993; Moss and Tyler, 1995). While there is general agreement that properties of concrete concepts include perceptual and functional features, the literature is less consistent about what exactly qualifies as a property of an abstract concept. When participants are specifically instructed to produce properties that they feel are characteristic of the concept itself, abstract concepts elicit fewer properties than concrete concepts (de Mornay Davies and Funnell, 2000; Tyler et al., 2002). Other studies with a broader definition of what qualifies as a property have found that concrete concepts elicit more properties that explicitly describe the concept (Barsalou and Wiemer-Hastings, 2005; Wiemer-Hastings and Xu, 2005), but have noted that the definition of a property can be extended to include persons, objects, and other elements of situations associated with the concept, as well as internal states and other meaning-bearing utterances. For example, the protocol used by Wiemer-Hastings and Xu classifies the words *good* and *want* in a participant's description of *HOPE* ("something will happen good, you really want something to happen," p. 736) as words that carry information about internal states ("introspective features"), and many elements of situations were observed in descriptions of abstract concepts in the present study, including mentions of persons (*DANGER* → "a policeman may face this in his job"), objects (*SUCCESS* → "great house"), and events (*MISCHIEF* → "crimes at night"). When information of this sort is not ignored, apparent differences in richness between concrete and abstract concepts disappear or become far less extreme (Wiemer-Hastings and Xu, 2005).

While the present research does tally the number of properties for each concept according to both broad and a narrow criteria, our primary motivation was not to determine whether concrete words possess more properties than abstract ones. Rather, the primary goal was to determine whether the descriptions elicited by abstract words in property generation tasks add to their richness in in a comparable manner to concrete words (i.e., whether "properties" of abstract concepts contribute to NF effects), and if so, what kinds of properties are most facilitative.

On some accounts, the situation-relevant and introspective utterances that participants use to describe abstract concepts in feature generation tasks are conceived of as properties in a strong sense, playing a central role in abstract concept representations (Barsalou and Wiemer-Hastings, 2005; Barsalou et al., 2008). If this is the case, one might expect that the quantity of introspective and situation properties that an abstract word elicits would predict its ease of processing, just as the number of perceptual properties does for concrete words (Grondin et al., 2009).

However, such utterances may not describe core components of the concept's representation at all. One possibility is that the words that participants use to describe abstract concepts are analogous to *associates*, i.e., participants' responses in free-association tasks. Studies of concrete concepts that have directly compared the influence of NF and *number of associates* (NoA) have shown the latter to have a relatively weak or undetectable impact (Yap et al., 2011; Rabovsky et al., 2012), and the same may be true for abstract concepts. A second possibility is that the words that participants use to describe abstract concepts may facilitate processing to the degree that they occur in similar linguistic contexts. It has been argued that language-based information plays a more important role in abstract (vs. concrete) concept representations (e.g., Sabsevitz et al., 2005; Borghi et al., 2011). NSN measures the number of words (NW) that occur in similar lexical contexts (Pexman et al., 2008), as approximated by counting the NW that occur within a particular radius of a high-dimensional semantic space. Such language-based measures of richness have been shown to predict LDT among concrete concepts (Buchanan et al., 2001; Pexman et al., 2008; Yap et al., 2011). According to theories that emphasize the importance of linguistic information for abstract concepts, such measures of the richness of a word's linguistic contexts should be even more predictive of processing differences among abstract stimuli.

Of course, this need not be framed as a dichotomy. For example, Kiefer and Pulvermüller (2011) argue that multimodal information from perception and action constitutes the core content of abstract concept representations, but also note that abstract concepts may be more dependent upon on word-based associations than concrete concepts. Furthermore, a measure of semantic richness is not wholly language-based merely because it is derived from a text corpus; word pairs that are highly related according to corpus-based measures such as LSA frequently refer to objects that occur together in the world (Baroni and Lenci, 2008; Louwerse, 2008). Even so, the fact that measures based on corpora and feature norms account for unique variance in lexical and semantic decision tasks suggests that corpus-based measures contain some information about associations between words as they are used in language that feature norms do not, and vice versa (see Riordan and Jones, 2011). Contrasting the predictive power of multiple measures of richness can thus inform our understanding of the relative importance of different types of information to concrete and abstract conceptual representations.

MEASURES OF SEMANTIC RICHNESS

The three variables considered by Pexman et al. (2008)—NF, CD, and NSN—are not the only ones that have been investigated as measures of semantic richness. Yap et al. (2011) extended this work in several ways. First, they included additional variables that had been proposed in the literature as indicators of semantic richness: NoA (Duñabeitia et al., 2008) in the Nelson et al. (1998) free-association norms, and *lexical ambiguity*, which they operationalized as a word's log-transformed number of senses in WordNet (Miller, 1990). Second, they used alternative CD and neighborhood measures that had been calculated on larger corpora and accounted for more variance than previous operationalizations of CD and NSN.

Finally, they included additional lexical control variables known to account for substantial variance in lexical decision and naming times (NTs). Using these measures, they found that neighborhood density, CD, and NF all accounted for unique variance above and beyond that accounted for by lexical-level variables in the lexical decision task, whereas number of senses and lexical ambiguity did not. Yap et al. also found that CD and NF, but not NSN, predicted unique variance in speeded pronunciation times, although the effects were less robust. This is consistent with previous findings of facilitation for words with many features (Ashcraft, 1978; Pexman et al., 2003), contexts (Adelman et al., 2006; Jones et al., 2012), and semantic neighbors (Buchanan et al., 2001; Shaoul and Westbury, 2010).

COLLECTING SEMANTIC RICHNESS NORMS FOR CONCRETE AND ABSTRACT WORDS

Our norming study builds on Yap et al. (2011) by replicating their pattern of effects for lexical decision and speeded pronunciation on a new set of stimuli containing words that vary widely in concreteness. The absence of publicly available feature norms for English abstract concepts required us to collect a large set of property generation norms. In a property generation study, participants describe the properties of a concept verbally or in writing. For example, presented with the concept *dog*, participants might produce descriptions such as *has four legs, is furry,* etc. This method has a long history of use by researchers wishing to gain insight into the representations of concrete concepts and categories (e.g., Rosch and Mervis, 1975; Hampton, 1979; McRae et al., 2005; Vinson and Vigliocco, 2008), and less frequently, events and abstract concepts (e.g., Barsalou and Wiemer-Hastings, 2005; Wiemer-Hastings and Xu, 2005; Vinson and Vigliocco, 2008). Property norms should not be interpreted as a verbatim readout of semantic representations (Medin, 1989), but rather as a reflection of systematic regularities in the ways that participants describe concepts. They can provide insight into a concept's underlying semantic representation, although not all aspects of meaning are equally well represented. Some aspects of a concept's representation are not easily verbalized, while others may be particularly salient due to their distinctiveness. This poses philosophical challenges to traditional views that interpret features as fundamental components of semantic representations. However, it is less problematic for positions that treat features as offering a window into aspects of semantic meaning (McRae et al., 2005) or as *ad-hoc* descriptions of perceptual simulations (Barsalou and Wiemer-Hastings, 2005). McRae et al. (2005) note that although the absence of biological features, internal features, etc., is "occasionally interpreted as a weakness of such norms, it may actually be a strength, because it appears that these general features play only a small role in object identification, language comprehension, and language production precisely because they are not salient and are true of large numbers of concepts" (p. 549). Overall, the impressive ability of measures based on feature norms to account for variability across a wide range of lexical processing tasks (e.g., Yap et al., 2012) attests to their utility in capturing important aspects of meaning.

Following the basic design of McRae et al. (2005), our participants completed a property generation task in the form of an online game within our *Semantic Pictionary* platform (Kievit-Kylar and Jones, 2011). Online "games with a purpose" (von Ahn, 2006) are becoming more commonplace in cognitive science to crowdsource information about the properties of common objects from Internet users (e.g., Singh et al., 2002; Speer et al., 2010). This method permitted participants to describe abstract words without constraining them to produce properties in the form of predicates such as *has wings, is fast,* etc. There is a high correlation between *ease of predication* (participants' ratings of how easy it is to put words into simple factual statements) and concreteness (Jones, 1985; de Mornay Davies and Funnell, 2000), leading some to surmise that abstract words have far fewer properties than concrete words (e.g., Plaut and Shallice, 1993). If a property is defined as a predicate, this is a foregone conclusion.

However, predicates are not the basic units of semantic analysis, but are rather only one way of expressing underlying semantic relationships. *Wing*, *passenger*, and *pilot* are all concepts that possess a meaningful semantic relationship to *airplane*, and language affords us an easy way to express these relationships as predicates (airplanes *have* wings, airplanes *carry* passengers, airplanes *need* pilots). However, *courthouse*, *crime*, and *justice* are all concepts that possess a meaningful semantic relationship to *law*, and may play a role in its semantic representation, even if it is difficult to produce a three- or four-word sentence that encapsulates the nature of this relationship. We risk missing important insights about the nature of abstract concept representation if we exclude such concepts from analysis simply because participants do not express them as predicates. On the other hand, if we interpret all frequent responses as "features," we risk being too inclusive. We do not pretend to have a solution to this dilemma, and believe there may be value in both broad and constrained notions of what constitutes a property. For this reason, we restricted our definition of NF to the number of {concept $\rightarrow$ word} pairs that matched a subset of predefined semantic relations identified in the literature as being of likely importance to concrete and abstract concept representations. However, we also created an additional variable, NW, which is simply a count of all words produced by at least six of the 30 subjects who generated descriptions for that word. Details on how each of these measures was calculated appear in the Methods section.

METHODS

Participants

Seven hundred and sixty six participants (57% female) participated via the Indiana University Psychology Department subject pool for partial course credit. An additional 208 participants recruited via Amazon Mechanical Turk who completed the study for a payment of $1 per session. All participants resided in the United States and reported English as their first language.

Materials

After surveying the literature on feature generation studies and abstract word representation, 593 English nouns were selected to be normed. Items used in the feature generation studies of Barsalou and Wiemer-Hastings (2005), McRae et al. (2005), Wiemer-Hastings and Xu (2005), and Vinson and Vigliocco (2008) were selected to facilitate comparison between the data to be collected and that collected by other researchers, to build upon previous findings that used existing datasets, and because these items were originally selected to represent a broad range of stimuli used in the semantic memory literature. Additional stimuli were selected from the MRC Psycholinguistic Database (Coltheart, 1981) in order to ensure our stimuli included words with a high level of diversity in frequency, length, and concreteness. All words were classified for concreteness/abstractness on the basis of their rated MRC concreteness (see Analysis 2). See Recchia and Jones (2012) for the complete set of stimuli.

Procedure

Property generation. Participants were asked to participate in an online game in which they would be required to describe various words, and were informed that a future participant would be responsible for guessing the words from their descriptions. Participants were asked to provide 10 short descriptive properties for each of 20 words that would help their partner guess the target word. Participants were instructed to describe the concept, not the word itself; i.e., clues about the letter that a word starts with or words that it rhymed with were not permitted. Participants were asked to fill in all 10 blanks, but the online application did not require all 10 blanks to be filled in order to move on to the next word. Word type was alternated for each participant (i.e., each concrete word followed an abstract word and vice versa), and word type that began the task was balanced across participants. While McRae et al. (2005) provided explicit instructions about the sorts of properties they wanted subjects to produce ("physical properties, such as internal and external parts, and how it looks, sounds, smells, feels, or tastes; functional properties, such as what it is used for; where, when and by whom it is used; things that the concept is related to, such as the category that it belongs in; and other facts, such as how it behaves, or where it comes from," McRae et al., p. 556), our instructions left this considerably more open-ended, asking participants to "provide 10 short descriptive properties for each word that will help your partner guess your noun, without specifically telling your partner which word you have," with further instructions emphasizing that participants were responsible for describing the concept, not surface features of the word itself (e.g., "rhymes with"). At the completion of the study, 93% of the original 593 stimuli had been described by at least 30 participants, the same number of participants per word recruited by McRae et al. (2005); words for which this was not the case were excluded from analysis. The resulting set of 550 words included 281 items from the McRae et al. norms.

Measures of semantic richness. Four measures of semantic richness were obtained for each of the 550 cue words: NW, NF, NSN, and CD. Each concept's NW was determined by counting the number of unique words (types) produced by at least six[1] of the 30 subjects who generated descriptions for that concept. The set of words produced by at least six participants in response to a given concept were reformatted as a list of {concept → word} pairs. Some pairs exhibited a clear semantic relationship that could be expressed as a predicate (e.g., {key → metal}: *is made of*), while others exhibited semantic relationships that were not necessarily expressed as predicates but were captured by Wu and Barsalou's (2009) taxonomy of semantic codes for generated properties (e.g., {danger → emergency}: *event*). Yet others matched categories not covered in the Wu and Barsalou taxonomy, but which have been hypothesized to be of particular importance to abstract concept representations, such as communicative acts and social institutions/artifacts (Barsalou and Wiemer-Hastings, 2005; Wiemer-Hastings and Xu, 2005; Borghi and Cimatti, 2009).

Although Wu and Barsalou (2009) reported high levels of rater agreement for concrete words, our initial attempts at using their taxonomy for our set of abstract words proved relatively unreliable, as did our initial attempts to use taxonomies developed for coding free-response protocols (Barsalou and Wiemer-Hastings, 2005; Wiemer-Hastings and Xu, 2005). After multiple rounds of classification of a subset of properties by two raters, we ultimately settled on the partial taxonomy detailed in Recchia and Jones (2012). It is not intended to represent a complete set of feature types, as it omits several property types hypothesized to be highly relevant to concrete word representations (e.g., functions; agentive actions; category coordinates). The primary reason for this was that we wished to include only those feature types for which high levels of rater agreement could be achieved. Are *bowl* and *spoon* best conceived of as *category coordinates* (e.g., eating utensils) or *associated entities*? Systematic disagreements of this nature between raters were generally solved either by collapsing multiple categories into one or omitting a category entirely. However, when fine-grained distinctions could be preserved while retaining high reliability, we generally did so. Agreement between two raters on a 500-feature subset of the data was quite good (Cohen's $\kappa = 0.78$), and so the remainder of the data was coded by a single rater. Each {concept, word} pair that was produced by at least six participants and which matched one of these codes was considered a feature and was included in the NF measure.

A common trade-off in the development of a coding scheme is between reliability and comprehensiveness; our criteria clearly lean toward reliability. Our exclusion of some feature categories means that some valid features will have escaped our NF measure, but the feature categories that are coded for are consistent between raters. Thus, our NW and NF measures represent broad and narrow ends of the spectrum of definitions of what constitutes a "feature." As with any measure of NF, it is important to keep in mind that exactly how features are defined is critically important to the interpretation of NF effects (or the absence thereof).

Finally, NSN was calculated for each concept. Pexman et al. (2008) used global semantic neighborhood values calculated by Durda et al.'s (2006) *WordMine2* application. According to this measure, a word's neighborhood consisted of all words occurring within a specific radius of the high-dimensional space defined by HAL (Lund and Burgess, 1996), a co-occurrence-based model of lexical semantics. Yap et al. (2011) replaced this with an alternative measure of corpus-based neighborhood density that reflected the mean cosine between a word and its 5000 closest neighbors in a HAL-like semantic space (Shaoul and Westbury, 2010). In both studies, while high NSN facilitated performance in lexical decision, NSN had null effects on semantic decision tasks. Each

[1]This was the same threshold used by Vinson and Vigliocco (Andrews et al., 2009), and is nearly identical to the 5-subject threshold used by McRae et al. (2005).

study noted that this was perhaps due to the fact that neighborhood measures conflate close and distant neighbors, which have opposite effects on processing in some circumstances (Mirman and Magnuson, 2008). Yap et al. (2011) suggested that this could be partially addressed by parametrically manipulating the number of neighbors considered. In addition, the size of the window used for assessing whether the nearby appearance of two words counts as a "co-occurrence" should be treated as a parameter that must be optimized (Bullinaria and Levy, 2007).

Rather than compute a definition of NSN theoretically tied to a vector space model such as HAL, we calculated NSN using pointwise mutual information (PMI), a measure of association frequently used in computational linguistics to contrast the actual probability of observing two items together (e.g., in the same window of text) with the probability of having observed them together if they had been independently distributed (Manning and Schütze, 1999). PMI is calculated as

$$\log_2 \frac{p(xy)}{P(x)P(y)} \tag{1}$$

where $P(x)$ represents the probability of observing word x if a random window of text is selected from the corpus, $P(y)$ the probability of observing word y, and $P(xy)$ the probability of observing x and y together. PMI has been shown to be a good predictor of human semantic similarity and synonymy judgments (Recchia and Jones, 2009), and allows for a straightforward manner of calculating a measure of NSN not tied to any particular semantic space model: one can simply count the NW having a PMI exceeding some threshold t, using a window size of w. Exploratory manipulation of these parameters indicated that using the TASA corpus (Zeno et al., 1995), a window size of 8 and a threshold of 7 maximized correlations between this measure of NSN and LDT, and that the same parameters maximized correlations to NTs as well. These were therefore the parameters used for the calculation of NSN in the analyses reported here.

Finally, following Yap et al. (2011), a measure of CD was obtained from the English Lexicon Project (Balota et al., 2007), and the number of senses attributed to each word by lexicographers was obtained from WordNet (Miller, 1990). What is referred to in the following analyses as CD refers specifically to log SUBTL-CD, the logarithm of the number of transcribed film and television programs in the SUBTLEX corpus (Brysbaert and New, 2009) in which a word appears.

EXPLORING THE SEMANTIC RICHNESS NORMS

In this section, we conduct three basic analyses to explore the effects of semantic richness in our norms on lexical decision and naming data extracted from Balota et al. (2007). First, we attempt to replicate the overall findings of Yap et al. (2011) using our larger norms that have greater variability in concreteness. Second, we repeat the analysis separately for the abstract and most concrete sets of words in our norms to evaluate whether facilitative effects of semantic richness are consistent across the two words types. Finally, we expand our NF variable into counts of particular *types* of features to determine which features are most responsible for explaining the variance in response latency, and whether these responsible feature types differ between concrete and abstract words.

ANALYSIS 1: REPLICATING Yap et al. (2011)

To attempt to replicate the effects obtained by Yap et al. (2011) with our dataset and our measures of NF and NSN, we conducted a hierarchical regression analysis to assess the impact of measures of semantic richness on lexical decision and NTs. We used a near-identical set of control variables to Yap et al., but omitted Coltheart's N and its analog for phonological neighbors, as these are measures designed to account for the same underlying construct as the improved Levenshtein distance measures (orthographic/phonological density). Thus, the control variables entered into the regression were log-frequency (SUBTLEX corpus), number of morphemes, number of syllables, orthographic Levenshtein distance 20 (Yarkoni et al., 2008), and phonological Levenshtein distance 20. To control for phonetic biases in voice key response time measurements, a set of dichotomous onset variables taking on values of 0 or 1 for each stimulus were used to code for the absence/presence of 13 phonetic features (Balota et al., 2007; Yap et al., 2011, 2012); these were entered as additional predictors in the regression analysis of NT latencies only.

The measures of semantic richness entered were NW, NF, NSN, and CD, described in the preceding section[2]. Z-scores of LDT and NTs were obtained from the English Lexicon Project (Balota et al., 2007). Control variables were also obtained from this dataset; these were entered in the first step of the regression, while measures of semantic richness were entered in the second step. Descriptive statistics for each of these variables are presented in **Table 1**. All 550 stimulus items were included in the regression.

Results

As anticipated, NF, NSN, and CD were each found to be independent predictors of variance in LDT ($p < 0.001$) even after variance from control variables had been accounted for. **Table 2** reports correlations between each pair of variables, while **Table 3** reports betas and p-values from the regression analysis. Consistent with Yap et al. (2011), CD remained a significant predictor for NTs while NSN dropped out[3], although we did not find our measure of NF to predict NTs.

[2]Like Yap et al. (2011, 2012), we did not include NoA as a predictor in our primary regression analyses due to the fact that NoA counts were available for only some of our stimuli, as this would have reduced the power of our analyses substantially. When the regressions reported in this paper were repeated with NoA included as a predictor, NoA was not significant in any analysis.

[3]Also consistent with Yap et al. was a failure to find any contribution of orthographic and phonological Levenshtein distance, most likely because of the overlapping variance accounted for by these two predictors. Indeed, the high correlation between orthographic and phonological Levenshtein distance gave the set of control variables a maximum VIF (variance inflation factor) of 8.3, indicating concerning levels of multicollinearity. However, omitting either (or both) of these predictors did not change which semantic richness variables predicted significant levels of variance; the only change in the pattern for control variables was that log-frequency became a significant predictor of lexical decision times ($p < 0.01$), a finding more consistent with the known influence of frequency on lexical decision times.

Table 1 | Descriptive statistics for stimulus characteristics (predictors and dependent variables).

	All 550 stimulus items	Set of 147 abstract items	Set of 147 most concrete items
	M (SD)	*M* (SD)	*M* (SD)
Rated concreteness (MRC norms)	5.07 (1.37)	3.21 (0.49)	6.24 (0.11)
CONTROL VARIABLES			
Log frequency (Brysbaert and New, 2009)	2.75 (0.76)	3.15 (0.82)	2.86 (0.63)
Number of morphemes	1.27 (0.53)	1.48 (0.68)	1.09 (0.31)
Number of syllables	1.78 (0.84)	2.03 (0.98)	1.60 (0.76)
Number of letters (length)	5.85 (1.97)	6.35 (2.09)	5.29 (1.80)
OLD20 (Yarkoni et al., 2008)	2.12 (0.84)	2.18 (0.68)	1.97 (0.79)
PLD20 (Yarkoni et al., 2008)	1.96 (0.94)	2.05 (0.84)	1.76 (0.88)
SEMANTIC RICHNESS VARIABLES			
Number of semantic neighbors	150.61 (97.77)	199.59 (91.00)	167.08 (96.40)
Number of features (Analysis 1)	11.08 (3.83)	8.56 (3.78)	13.00 (3.24)
Number of features (McRae et al., 2005)	14.13 (3.64)	–	15.40 (3.59)
Number of words (Analysis 1)	32.10 (7.47)	27.36 (7.26)	35.61 (6.33)
Log ctx. dispersion (Brysbaert and New, 2009)	2.51 (0.68)	2.91 (0.67)	2.59 (0.57)
Log number of senses (Miller, 1990)	1.32 (0.82)	1.57 (0.79)	1.26 (0.75)
DEPENDENT MEASURES			
Lexical decision task RT (Balota et al., 2007)	653.16 (84.53)	644.35 (78.55)	630.71 (74.49)
Lexical decision task RT, standardized	−0.48 (0.28)	−0.52 (0.27)	−0.56 (0.26)
Pronunciation task RT (Balota et al., 2007)	635.37 (61.17)	634.23 (58.83)	624.97 (49.73)
Pronunciation task RT, standardized	−0.42 (0.26)	−0.42 (0.26)	−0.47 (0.22)

Note: OLD20, Orthographic Levenshtein Distance 20, a measure of orthographic neighborhood density; PLD20, Phonological Levenshtein Distance 20, a measure of phonological neighborhood density. The number of features measure in the McRae et al. (2005) dataset limited to those concepts that were used as stimuli in our dataset and which were also members of the McRae norms (first column, N = 281; third column, N = 93). Average rated concreteness for the entire stimulus set was computed over only the 446 words for which MRC concreteness ratings were available.

Discussion

Despite differences in our stimuli and in our measures for NF and NSN, we found a pattern of effects consistent with those reported elsewhere in the literature—particularly for lexical decision—giving us confidence that our measures tap into semantic richness constructs similar to those investigated by other researchers. However, this in itself tells us nothing about whether different types of richness contribute differentially to the processing of abstract vs. concrete concepts, as the majority of our dataset consisted of concrete concepts. In Analysis 2, we examine whether the same pattern of effects holds for the most abstract and most concrete words in our dataset.

ANALYSIS 2: SEMANTIC RICHNESS PREDICTIONS FOR CONCRETE vs. ABSTRACT WORDS

Consistent with prior research, Analysis 1 found unique contributions of NF, NSN, and CD in lexical decision. Do we observe differential effects for concrete and abstract words? If NSN and NF pattern differently for words of different levels of concreteness, this would lend support to theories that predict differences in the involvement of language in abstract and concrete representations.

Two regressions were conducted using the same methods described in Analysis 1, but were restricted to the abstract stimuli and an equally sized subset of the most concrete stimuli rather than the entire set (Recchia and Jones, 2012). MRC concreteness ratings were available for 446 of the 550 stimuli used in the feature generation task. Of these, 147 met the criteria used by Wiemer-Hastings and Xu (2005) to define abstractness (MRC rating lower than 4.5; see Wiemer-Hastings and Xu, 2005, Appendix A). This set of 147 abstract concepts was contrasted with a set of the 147 most concrete concepts (stimuli with the highest MRC ratings). **Table 1** provides descriptive statistics for the dataset as a whole and for the abstract/concrete subsets. Hierarchical regressions were computed with the same sets of control and semantic richness variables as in the previous analysis.

Results

Table 4 reports betas and p-values for the variables predicting lexical decision and naming latencies. For the set of 147 abstract words, NSN and CD were found to be significant predictors of LDT ($p < 0.05$), but NF was not ($p = 0.15$). In contrast, for the set of the 147 most concrete items, NF ($p < 0.01$) and CD ($p < 0.001$) were found to be significant predictors of LDT, but NSN was not ($p = 0.14$). When LDT regressions were repeated with one or neither Levenshtein distance predictor, there were no changes in which semantic richness variables remained significant and non-significant predictors. With Levenshtein variables omitted, variance inflation factors (VIFs) were acceptably low for both abstract concepts (4.0 and 1.8 for control and semantic variables, respectively) and concrete concepts (3.3 and 3.0).

Table 2 | Intercorrelations among predictor and dependent variables in the regression analyses for all stimuli.

	NSN	NF_{RJ}	NF_M†	NW	CD	Freq	Nm	Ns	Len	OLD	PLD	Sens	LDT_{raw}	LDT_Z	NT_{raw}	NT_Z
NSN	–	−0.199**	0.189**	0.055	0.729**	0.686**	−0.035	−0.191**	−0.201**	−0.311**	−0.307**	0.496**	−0.519**	−0.551**	−0.366**	−0.373**
NF_{RJ}		–	0.238**	0.475**	−0.169**	−0.133**	−0.163**	−0.031	−0.057	0.068	0.051	−0.307**	0.009	0.011	−0.004	0.013
NF_M†			–	0.098	0.219**	0.232**	0.057	−0.093	−0.089	−0.081	−0.089	0.037	−0.235**	−0.243**	−0.102	−0.119*
NW				–	0.095*	0.093*	−0.150**	−0.139**	−0.131**	−0.075	−0.083	0.005	−0.113**	−0.129**	−0.130**	−0.124**
CD					–	0.985**	−0.157**	−0.314**	−0.354**	−0.435**	−0.422**	0.582**	−0.634**	−0.685**	−0.507**	−0.514**
Freq						–	−0.195**	−0.333**	−0.379**	−0.442**	−0.430**	0.562**	−0.623**	−0.673**	−0.496**	−0.504**
Nm							–	0.541**	0.621**	0.465**	0.502**	−0.230**	0.291**	0.323**	0.305**	0.291**
Ns								–	0.820**	0.756**	0.804**	−0.386**	0.506**	0.527**	0.466**	0.451**
Len									–	0.871**	0.852**	−0.375**	0.559**	0.598**	0.551**	0.558**
OLD										–	0.916**	−0.457**	0.570**	0.607**	0.510**	0.519**
PLD											–	−0.442**	0.547**	0.587**	0.499**	0.511**
Sens												–	–	–	–	–
LDT_{raw}													–	0.949**	0.647**	0.642**
LDT_Z														–	0.684**	0.682**
NT_{raw}															–	0.948**
NT_Z																–

Note: NSN, number of semantic neighbors; NF_{RJ}, number of features, Analysis 1; NF_M, number of features, McRae et al. (2005); NW, number of words; CD, log contextual dispersion; WF, log word frequency; Nm, number of morphemes; Ns, number of syllables; Len, number of letters; OLD, Orthographic Levenshtein distance 20; PLD, Phonological Levenshtein distance 20; Sens, log number of senses; LDT_{raw}, lexical decision time; LDT_Z, lexical decision time (z-scored); NT_{raw}, naming time; NT_Z, naming time (z-scored).

†NF_M has a large number of missing values due to the fact that only 281 words in the McRae et al. (2005) norms also appear in the present dataset.

$**p < 0.01$; $*p < 0.05$.

Table 3 | Standardized regression coefficients predicting lexical decision and naming latencies, using number-of-features measure derived from data collected in Analysis 1.

Variables	Betas	
	LDT	NT
STEP 1: ONSETS		
Adjusted R^2	0.00	0.17***
STEP 2: CONTROL VARIABLES		
Log frequency	−0.505	−0.352***
Number of morphemes	−0.036	−0.065†
Number of syllables	0.071	0.059
Number of letters (length)	0.286***	0.407***
OLD20	0.105	−0.075
PLD20	−0.009	0.060
Adjusted R^2	0.59	0.55
Change in R^2	0.59***	0.38***
STEP 3: SEMANTIC RICHNESS VARIABLES		
Number of features	−0.114***	−0.036
Number of words	0.027	−0.021
Number of semantic neighbors	−0.135***	−0.027
Log contextual dispersion	−0.848***	−1.038***
Log number of senses	−0.051	0.034
Adjusted R^2	0.63	0.58
Change in R^2	0.04***	0.03***

Note: LDT, lexical decision time (z-scored); NT, naming time (z-scored); OLD20, Orthographic Levenshtein Distance 20, a measure of orthographic neighborhood density; PLD20, Phonological Levenshtein Distance 20, a measure of phonological neighborhood density. Only semantic richness variables are shown in Step 2 for ease of exposition.

$^\dagger p < 0.10$; $***p < 0.001$.

We considered the possibility that NSN was not a significant predictor for the most concrete words merely because the variance accounted for by NF and NSN overlapped in such a way that NSN would have been a significant predictor for the most concrete words had NF not been part of the analysis. However, NSN was still not a significant predictor for the set of the most concrete words even when NF was omitted from the regression ($p = 0.2$). Similarly, NF was not a significant predictor for the set of abstract words even when NSN was omitted from the regression ($p = 0.4$).

For NTs, CD was the only significant semantic richness predictor for abstract and concrete concepts. NF was a marginally significant predictor for concrete concepts ($p = 0.07$), but not for abstract concepts ($p = 0.28$).

Discussion

As previously described, different theories of concept representation make different predictions with respect to the role of language and semantic features for abstract concepts. Internal experiences (felt experiences of judgments, cognitive operations, emotional valence, etc.) have been hypothesized to play a special role in grounding abstract concepts, as have complex situations involving multiple actors, particularly social actors (Barsalou and Wiemer-Hastings, 2005; Wiemer-Hastings and Xu, 2005; Borghi and Cimatti, 2009). Indeed, the set of abstract concepts under investigation was rich in these categories of features. Abstract concepts were relatively high in several categories of features hypothesized by these researchers to be of particular importance, with an average NF per concept of 0.66 for communicative acts (vs. 0.07 for concrete concepts), 0.61 for evaluations (vs. 0.34 for concrete concepts), 1.35 for social artifacts/actions (vs. 0.26 for concrete concepts), and 2.6 for cognitive states/operations/affects—a higher mean than in any single feature category for the most concrete concepts, although concrete concepts elicited more features overall.

The fact that abstract concepts were so frequently described in terms of internal and social experiences hints that these may indeed be important aspects of abstract concept representation. However, the present analyses suggest that being rich in these kinds of features likely does not facilitate early processing of abstract concepts in the same way that being feature-rich facilitates early processing of highly concrete concepts. Of course, it is certainly possible that annotating participants' descriptions for other kinds of features would yield different results. It is also possible that 147 abstract concepts was simply too few to detect a NF effect. However, the fact that NF was a significant predictor of the 147 most concrete concepts' LDT ($p < 0.01$), and approached significance in the NT regression[4] ($p = 0.07$), implies a stronger role for features in concrete concept representations.

Another way in which our results differed between abstract and concrete concepts was in the degree of facilitation provided by NSN. In cross-task comparisons, semantic neighborhood density has been shown to facilitate concrete concept processing in lexical decision, but not in other tasks such as semantic decision or word naming (Yap et al., 2012). Analysis 1 replicated this pattern of results: Using a large dataset consisting of primarily concrete concepts, NSN was found to be a significant predictor of lexical decision but not NTs. In Analysis 2, however, no effect of NSN was found on a smaller dataset consisting of only the most concrete concepts. This may have been due to the loss of statistical power resulting from the smaller subset of stimuli (147 items). However, the fact that NSN *was* a significant predictor for an equally small set of abstract concepts suggests an important dissociation.

Given that NSN represents the richness of the linguistic contexts in which words denoting particular concepts appear, a

[4]For the regression using naming times of concrete concepts as the dependent variable, two unexpected results were the extraordinarily high amount of variance accounted for by the phonetic onset variables (adjusted $r^2 = 0.25$) and the lack of significance for word length (**Table 4**). Closer inspection of the data suggested that this was likely the result of a chance correlation between word length and onset characteristics. Of the 13 dichotomous phonetic onset variables, the one having the strongest point-biserial correlation with word length was the *approximant* variable ($r_{pb} = -0.16$, $p = 0.06$); this variable also accounted for the most variance (compared with other onset variables) among naming times in the regression ($b = -0.46$, $p = 0.005$). In other words, our concrete stimuli that began with approximants happened to be particularly short words, which may have caused phonetic onsets to appear to be stronger predictors of variance than they truly were.

Table 4 | Standardized regression coefficients predicting lexical decision and naming latencies.

Variables	Betas			
	Set of 147 abstract items		Set of 147 most concrete items	
	LDT	NT	LDT	NT
STEP 1: ONSETS				
Adjusted R^2	0.00	0.14**	0.00	0.25***
STEP 2: CONTROL VARIABLES				
Log frequency	−0.413***	0.331***	−0.514***	−0.291***
Number of morphemes	0.094	−0.013	−0.080	−0.136*
Number of syllables	−0.004	−0.195	−0.098	0.025
Number of letters (length)	0.444**	0.575***	0.223†	0.234
OLD20	0.053	−0.231†	0.066	−0.186
PLD20	−0.119	0.309*	0.238	0.313†
Adjusted R^2	0.54	0.54	0.62	0.53
Change in R^2	0.54***	0.40***	0.62***	0.28***
STEP 3: SEMANTIC RICHNESS VARIABLES				
Number of features	−0.089	0.072	−0.168**	−0.115†
Number of words	0.024	0.008	0.083	0.006
Number of semantic neighbors	−0.147*	−0.072	−0.121	−0.060
Log contextual dispersion	−0.890*	−1.094*	−0.944***	−1.069**
Log number of senses	−0.043	0.045	−0.024	0.045
Adjusted R^2	0.59	0.58	0.67	0.57
Change in R^2	0.04***	0.04***	0.05***	0.04***

Note: LDT, lexical decision time (z-scored); NT, naming time (z-scored); OLD20, Orthographic Levenshtein Distance 20, a measure of orthographic neighborhood density; PLD20, Phonological Levenshtein Distance 20, a measure of phonological neighborhood density. Only semantic richness variables are shown in Step 2 for ease of exposition.

*†$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.*

greater role for NSN in abstract than concrete concept representations is consistent with hypotheses that abstract representations are more heavily grounded in language than are concrete representations, whether researchers attribute this to differences in the process by which abstract concept meanings are acquired (Della Rosa et al., 2010; Borghi et al., 2011) or to abstract concept representations' purported dearth of multimodal perceptual information (Plaut and Shallice, 1993; Paivio, 2010). This finding is not necessarily inconsistent with theories that ground both concrete and abstract concepts in nonlinguistic content (e.g., Barsalou et al., 2008), as long as these theories can be extended to explain why measures of language-based richness are more predictive of LDT for abstract words than for concrete ones. Theories in which abstract concepts are represented primarily as conceptual metaphors (Lakoff, 2009) would also require additional scaffolding to accommodate this result.

So far, these analyses do not tell us *which* features are facilitating the processing of concrete words, nor if there are subsets of features that may be differentially facilitating abstract word processing. For example, situation properties, social institutions/artifacts (Barsalou and Wiemer-Hastings, 2005), and internal experiences (Barsalou et al., 2008) have been argued to be central to abstract word representations, as have aspects of meaning related to social/communicative function (Borghi and Cimatti, 2009; Borghi et al., 2011). Does the number of such specific properties generated in a feature generation task predict LDT for abstract words? This issue is investigated in the final analysis.

ANALYSIS 3: WHAT TYPES OF FEATURES FACILITATE PROCESSING OF CONCRETE vs. ABSTRACT WORDS?

Although our composite NF variable did not predict lexical decision or NTs for the set of abstract concepts considered in Analysis 2, perhaps the number of particular kinds of features would have proved to be reliable predictors if they had been considered as separate variables. In addition, it seems likely that not all types of features are equally important for the NF effects observed for concrete concepts. Features representing differing knowledge types have been shown to follow different timecourses of activation (Amsel, 2011), some of which are protracted enough that it is unlikely that they would have an influence on lexical decision. The purpose of Analysis 3 was to investigate what types of features account for the lion's share of the variance predicted by NF, and whether this differed between abstract and concrete concepts.

ANALYSIS 3a: FINE-GRAINED SEMANTIC CATEGORIES

NF was decomposed into 19 separate variables, each of which represented the NFs of a particular type (e.g., number of locations, number of visual properties, etc.) detailed in Recchia and Jones (2012), and the regressions from Analyses 1 and 2 were repeated. Descriptive statistics on the NFs in each category are displayed in **Table 5**. Three separate regressions were again conducted on all stimuli, on abstract stimuli only, and on most-concrete stimuli only, using the same methods as in Analyses 1 and 2, but with NF broken into 19 separate predictors representing the NFs of different semantic types. VIFs for the set of semantic richness variables remained low after these transformations, with max VIFs of 2.9, 2.0, and 3.8 for the set of all stimuli, abstract stimuli, and most-concrete stimuli, respectively, and VIFs of each of the 19 new predictors less than 1.7. Furthermore, because only LDT showed a reliable effect of NF in Analyses 1 and 2, NTs were not included as a dependent measure.

Although we coded our own data according to our own feature taxonomy, we also wished to take advantage of the fact that the McRae et al. (2005) feature norms have been similarly annotated with a fine-grained set of semantic relations. Specifically, the WB_Label field labels each of McRae et al.'s 7259 {concept, feature} pairs with one of 27 categories nearly identical to those described in Wu and Barsalou (2009, Appendix A). Because these norms constitute a separate dataset to which a separate set of feature codes has been applied by other raters, we hoped that including a comparable analysis that utilized this dataset might offer complementary insights as to which feature types most strongly drive NF effects. We therefore also repeated the regression conducted in Analysis 1 on the subset of 281 concepts that occurred in the McRae et al. (2005) norms as well as our own stimuli, replacing NF with 27 separate variables, each of which represented the NFs in the McRae et al. norms of a particular WB_Label type (i.e., internal components, locations, and the 25 other feature types appearing in their WB_Label column). These were not highly intercorrelated, with semantic richness variables exhibiting a max VIF of 4.1, and the 27 new variables' VIFs being less than 2.0 in all cases.

Results

For the regression conducted on the set of 147 abstract words, none of the 19 new NF variables accounted for unique variance in LDT. Similarly, for the set of 147 most concrete words, none of the 19 NF variables individually accounted for unique variance in LDT. However, for the full set (analog to Analysis 1), the number of the following kinds of features accounted for unique variance: *locations* (locations in which the concept is found; $p < 0.001$), *associated entities* (objects that tend to co-occur in real-world situations with the concept; $p < 0.05$), and *larger continuous wholes* (objects that are made out of a material described by the concept; $p < 0.05$). No other feature classes were a significant predictor of

Table 5 | Descriptive statistics for type counts of different feature categories.

	All 550 stimulus items	Set of 147 abstract items	Set of 147 most concrete items
	M (SD)	*M* (SD)	*M* (SD)
Num. communicative acts (com)	0.28 (1.00)	0.66 (1.31)	0.07 (0.48)
Num. materials (has_material)	0.49 (0.92)	0.01 (0.12)	0.77 (1.24)
Num. components (has_part)	1.17 (1.60)	0.04 (0.23)	1.67 (1.66)
Num. larger continuous wholes (is_material_of)	0.21 (0.76)	0.00 (0.00)	0.56 (1.23)
Num. larger discrete wholes (is_part_of)	0.08 (0.45)	0.01 (0.08)	0.16 (0.67)
Num. visual properties (vis)	0.50 (0.88)	0.09 (0.33)	0.73 (1.06)
Num. non-visual perceptual properties (perc)	1.35 (1.55)	0.05 (0.21)	2.07 (1.54)
Num. cognitive states/operations/affects (cog)	1.00 (1.98)	2.61 (3.05)	0.27 (0.51)
Num. contingencies (conting)	0.07 (0.29)	0.16 (0.42)	0.05 (0.24)
Num. evaluations (eval)	0.42 (0.84)	0.61 (1.16)	0.34 (0.66)
Num. negations (neg)	0.31 (0.52)	0.48 (0.63)	0.21 (0.44)
Num. social artifacts/actions (soc)	0.58 (1.35)	1.35 (2.01)	0.26 (0.59)
Num. events (ev)	0.23 (0.57)	0.19 (0.46)	0.27 (0.71)
Num. locations (loc)	0.85 (1.15)	0.29 (0.80)	1.20 (1.17)
Num. manners (man)	0.05 (0.23)	0.05 (0.21)	0.07 (0.26)
Num. participants (par)	0.64 (0.96)	0.78 (1.05)	0.62 (0.95)
Num. associated entities (ae)	1.48 (1.51)	0.65 (1.09)	1.71 (1.45)
Num. times (time)	0.31 (0.81)	0.39 (1.00)	0.21 (0.43)
Num. super/subordinates (tax)	1.06 (1.20)	0.14 (0.40)	1.75 (1.25)
SUPERCATEGORIES			
Num. entity properties	3.80 (3.30)	0.20 (0.51)	5.97 (2.74)
Num. introspective properties	1.79 (2.42)	3.86 (3.40)	0.86 (0.94)
Num. taxonomic properties	1.06 (1.20)	0.14 (0.40)	1.75 (1.25)
Num. concrete situation properties	2.33 (1.89)	0.95 (1.40)	2.91 (1.65)
Num. other situation properties	1.23 (1.47)	1.41 (1.59)	1.18 (1.30)

unique variance in our data. Betas and significance levels for this analysis are reported in Recchia and Jones (2012).

For the regression using the feature counts, semantic classes, and data from McRae et al. (2005), *locations* ($p < 0.05$) and *associated entities* ($p < 0.001$) were the only two variables that explained significant unique variance. No other feature classes accounted for unique variance in this dataset. The full set of betas and significance levels for this analysis are listed in Recchia and Jones (2012).

Discussion

Our replications of Analysis 2 with specific feature types proved inconclusive: When restricting the regressions to 147-concept subsets and dividing NF into 19 variables, none of the variables accounted for significant levels of variance in LDTs, although this could be due to data sparsity. However, on the two replications of Analysis 1 using our data and the data from McRae et al. (2005), there was converging evidence that NF effects for concrete words are primarily driven by *locations* and *associated entities*. Although both location and associated entity features were relatively common, their predictive power does not seem to derive solely from their overall high frequency. For example, several classes of properties (non-visual perceptual properties, components, subordinates/superordinates, cognitive states/operations) were more frequent than locations.

The finding that *locations* and *associated entities* were predictive in both datasets is especially striking, considering the substantial differences in the process according to which features were coded in each. As previously described, the NF measure in the McRae et al. norms is a count of the number of distinct predicates used to describe each concept, whereas our measure of NF was obtained for each concept by counting the number of distinct {concept → word} pairs that matched a set of predefined semantic relations. Although an investigation of three publicly available sets of properties (Howell et al., 2005; McRae et al., 2005; Vinson and Vigliocco, 2008) and two sets of associates (Kiss et al., 1973; Nelson et al., 1998) indicated that our concept vectors were more highly correlated with those of the McRae norms than were those of the other datasets (Recchia et al., 2011), the correlation between the *number of associated entity* variables in the two regressions conducted in Analysis 3a is weak ($r = 0.24$). This is likely due to differences in the way the category of *associated entity* was defined in our coding categories vs. the Wu and Barsalou (2009) categories used in the McRae norms (see General Discussion). Thus, one might reasonably argue that the *number of associated entity* variables in these regressions tap different constructs, although the fact that both are estimates of the number of distinct objects that occur together with the concept in real-world situations is suggestive. However, the *location* categories in the two coding systems have very similar definitions, and the correlation between the *number of location* variables is substantially higher ($r = 0.44$). The fact that *number of locations* accounts for unique variance across multiple datasets and coding schemas suggests that being associated with many physical contexts facilitates lexical decision latencies for concrete concepts.

Surprisingly, variables such as *number of visual properties* did not predict LDT, even though shared visual form/surface properties have predicted LDT in at least one previous study (Grondin et al., 2009). Our failure to detect an effect may have been due in part to the high fractionation in feature types, resulting in low statistical power. Our final analysis again replicates Analysis 1, but groups features into the four supercategories proposed by Wu and Barsalou (2009): entity properties (physical or systemic properties of the entity itself, such as visual properties), situation properties (properties of situations in which the entity occurs), introspective properties (properties of mental states associated with the concept), and taxonomic properties (hypernyms, hyponyms, etc.). The two feature types that predicted unique variance in Analysis 3 (locations and associated entities) are unique among situation properties in that they pick out *concrete objects*—i.e., they answer the question, "what things co-occur with this concept?" As such, they are distinguished from other situation properties in the following analysis.

ANALYSIS 3b: COARSE-GRAINED SEMANTIC CATEGORIES

All regressions performed in Analysis 3a were repeated, with two differences: First, the 27 NF variables corresponding to feature subtypes in the McRae norms were grouped into five supercategories representing the number of *entity properties*, *introspective properties*, *taxonomic properties*, *concrete situation properties* (associated entities and locations), and *other situation properties*. See Wu and Barsalou (2009) for the taxonomy of which subtypes belong in which supercategories. Second, the feature subtypes in our own norms were similarly reclassified according to their reference number in Recchia and Jones (2012) as *entity properties* (2, 3, 4, 5, 6, 7), *introspective properties* (8, 9, 10, 11), *taxonomic properties* (19), *concrete situation properties* (14, 17), and *other situation properties* (13, 15, 16, 18). Means and standard deviations for the NFs in each supercategory are reported in **Table 5**.

Results

In our own norms, the supercategory variables that predicted lexical decision latency were *number of entity properties* ($p < 0.05$), *number of concrete situation properties* ($p < 0.01$), and *number of other situation properties* ($p < 0.05$). Predictive supercategory variables for the McRae subset were *number of entity properties* ($p < 0.05$) and *number of concrete situation properties* ($p < 0.01$). The correlation between *number of entity properties* variables calculated using the McRae data/codes vs. our own data/codes was 0.41. **Tables 6** and **7** report betas and significance levels for these analyses.

As before, no supercategory variables predicted lexical decision latencies for the set of 147 abstract stimuli. In contrast, *number of entity properties* predicted lexical decision latencies for the set of the 147 most concrete nouns, $b = -0.167$, $p = 0.002$.

Discussion

As expected from the results of Analysis 3a, the number of *concrete situation properties* (associated entities and locations) was a strong predictor of lexical decision latency in both datasets. In addition, the number of *entity properties* facilitated lexical decision, a fact obscured by the many subcategories in

Table 6 | Standardized regression coefficients predicting lexical decision latencies, using feature counts and codes from data collected in Analysis 1.

Variables	Betas
STEP 1: CONTROL VARIABLES	
Log frequency	−0.505***
Number of morphemes	−0.036
Number of syllables	0.071
Number of letters (length)	0.286***
OLD20	0.105
PLD20	−0.009
Adjusted R^2	0.59
STEP 2: SEMANTIC RICHNESS VARIABLES	
Number of words	0.056†
Number of semantic neighbors	−0.130***
Log contextual dispersion	−0.940***
Log number of senses	−0.045
Num. entity properties	−0.074*
Num. introspective properties	−0.037
Num. taxonomic properties	−0.051
Num. concrete situation properties	−0.098**
Num. other situation properties	−0.060*
Adjusted R^2	0.63
Change in R^2	0.04***

Note: OLD20, Orthographic Levenshtein Distance 20, a measure of orthographic neighborhood density; PLD20, Phonological Levenshtein Distance 20, a measure of phonological neighborhood density. Only semantic richness variables are shown in Step 2 for ease of exposition.
*†$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.*

Table 7 | Standardized regression coefficients predicting lexical decision latencies, for all stimuli used in Analyses 1–2 that occur in the McRae et al. (2005) norms, using feature counts and codes from the McRae et al. (2005) dataset.

Variables	Betas
STEP 1: CONTROL VARIABLES	
Log frequency	−0.538***
Number of morphemes	−0.081†
Number of syllables	0.083
Number of letters (length)	0.299**
OLD20	−0.006
PLD20	0.051
Adjusted R^2	0.62
STEP 2: SEMANTIC RICHNESS VARIABLES	
Number of words	−0.007
Number of semantic neighbors	−0.021
Log contextual dispersion	−0.450†
Log number of senses	−0.075
Num. entity properties	−0.083*
Num. introspective properties	−0.002
Num. taxonomic properties	−0.005
Num. concrete situation properties	−0.101**
Num. other situation properties	0.017
Adjusted R^2	0.64
Change in R^2	0.02***

Note: OLD20, Orthographic Levenshtein Distance 20, a measure of orthographic neighborhood density; PLD20, Phonological Levenshtein Distance 20, a measure of phonological neighborhood density. Only semantic richness variables are shown in Step 2 for ease of exposition.
*†$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.*

the previous analysis. *Entity properties* are properties of the concept itself, both physical (visual, auditory, etc.) and systemic (e.g., the concept's components, or entities of which it is a component); they are extremely infrequent for abstract concepts–see the General Discussion and Recchia and Jones (2012), categories 2–7 for examples. One might argue that *number of entity properties* reached significance merely because it was the most frequent property type. However, *number of introspective properties* did not predict unique variance for the 147 abstract stimuli, despite the fact that introspective properties were more frequent for this group than *number of entity properties* were for the group of the most concrete stimuli (**Table 5**). In contrast, *number of entity properties* did predict unique variance for the most concrete stimuli ($p = 0.002$). Overall, *entity* and *concrete situation* properties appear to drive NF effects, consistent with the finding of Analysis 2 that NF predicted LDT for concrete but not abstract words.

GENERAL DISCUSSION

We replicated the general findings from Pexman et al. (2008) and Yap et al. (2011) that NSN, NF, and CD all account for unique variance in LDT. However, repeating this analysis for only the abstract words and for an equally sized subset of the most concrete words, NSN (but not NF) facilitated processing for abstract words, while NF (but not NSN) facilitated processing for the most concrete words. As Yap et al., (2011, 2012) noted, results for NTs were generally attenuated and did not show reliable effects of NF or NSN. However, length, log frequency, and CD generally predicted significant levels of variance for naming as well as LDT. Due to the correlation that Levenshtein distance measures (OLD20, PLD20) shared with each other and with log frequency, log frequency was not always a reliable predictor of LDT when these variables were included in the regression, but it was a consistent predictor when one or both of these variables were omitted.

With respect to the types of features that drive NF effects in lexical decision, we did not find evidence to suggest that having a high number of introspective, situation properties, social, or communication-related properties facilitated the processing of abstract or concrete words. This may have simply been due to data sparsity, as no fine-grained feature type predicted LDT in the analyses that were restricted to the 147 abstract or the 147 most concrete stimuli in the dataset. However, when analyzing our entire set of stimuli, we found a similar pattern of effects across two sets of feature norms—ours and those of McRae et al. (2005)—showing in each case, that the number of *entity* *properties* and *concrete situation properties* (*locations* and *associated entities*) that were attributed to a concept predicted its lexical decision latency. *Entity properties* generally refer to physical characteristics of objects, and are far more frequent for concrete words ($M = 3.80$, $SD = 3.30$) than for abstract words ($M = 0.20$, $SD = 0.51$). Generally speaking, entity properties were only attributed to abstract words that were capable of being visualized or audiated despite being rated as abstract (e.g., hell → **brimstone**; crash → **loud**), or which had structured components (story → **plot**). *Locations* are references to places in which the concept might be located; examples from our data include helicopter → **air**, raisin → **box**, pigeon → **city**, cancer → **lung**, etc. These, too, were quite frequent for concrete nouns (**Table 5**). Examples for abstract concepts were rare, but did occur, e.g., ache → **tooth**, thought → **head**, heaven → **clouds**, plea → **court**, etc.

Associated entities are similar to locations in that they pick out entities that co-occur with the concept in real-world situations. However, the category is significantly broader: an associated entity need not *be* the location in which the concept is located, but it may occur *in* the same location as the concept in the real-world. Examples include *squirrel* → ***acorns***, *beach* → ***castles***, etc. Note that in these particular examples, there is no similarity relation: squirrels are not similar to acorns, and beaches are not similar to castles. However, our raters found that in many cases, it was extremely difficult to disentangle these notions, as many items that are similar to each other also tend to occur in similar contexts (*comb* → ***brush***, *broccoli* → ***cauliflower***, *wall* → ***ceiling***, *spoon* → ***fork***, etc.). Therefore, we did not attempt to distinguish the two types of relation, but collapsed them both under a single category: as noted in Recchia and Jones (2012), we defined code 17 as "An object similar to the entity, or tending to co-occur with the concept in real-world situations." The definition of *associated entity* in the Wu and Barsalou coding scheme that was used to code the McRae norms is considerably narrower: "an entity in a situation that contains the focal concept" (Wu and Barsalou, 2009, p. 187). Despite these differing definitions, the number of associated entities per concept predicted LDT when analyzing both our data/codes and the data/codes of McRae et al.

Given that NF facilitated processing for the set of the 147 most concrete concepts, but not for the set of 147 abstract concepts, and that the full dataset consisted primarily of concrete concepts, it seems likely that being rich in locations, associated objects, and salient physical characteristics (entity properties) facilitates lexical decision for concrete concepts. It is inconclusive whether this is the case for abstract concepts. One intriguing similarity between locations and associated entities is that each picks out concrete entities (places or objects) that co-occur with the concept in day-to-day situations. This suggests that the features that facilitate concrete concept processing include those that pick out a concept's real-world contexts (cf. Hare et al., 2009). The consistency in the pattern of effects observed suggests that NF's ability to predict unique LDT variance owes at least in part to the fact that it captures the number of places and objects associated with a concept in the real-world.

In contrast, rich *linguistic* contexts (many semantic neighbors) appear to facilitate early activation of abstract concepts, as demonstrated in Analysis 2. This may be due to the fact that we acquire and use abstract words primarily in social situations in which language is highly salient (Borghi and Cimatti, 2009), or that we have no choice but to ground abstract words in language definitions because they have no perceptual referents (Paivio, 1986), or because language use encodes information about both abstract and concrete words (Louwerse, 2008) and the information so encoded is simply more relevant to abstract concept semantics. Given that NSN predicted LDTs for the entire dataset—composed primarily of concrete words—semantic density certainly seems to have a role to play for concrete concepts. However, Analysis 2 demonstrates that the *relative* influence of NF appears to be greater than that of NSN for the most concrete words, whereas the reverse appears to be true for abstract words. Furthermore, this dissociation does not seem to be an artifact of overlapping variance: For the most concrete words, NSN remained insignificant even when NF was removed from the regression, whereas for the most abstract words, NF remained insignificant even when NSN was removed from the regression. This was as expected, given the low correlation between NF and NSN.

While these results do not rule out the possibility that abstract words are simply grounded in different sorts of features than are concrete words, it appears that features of the kind we have investigated in this study do not provide the same sort of processing advantage for abstract as for concrete words. This is perhaps not surprising, given the shallowness of processing that is required for lexical decision—simulation of emotions, internal states, communication-related words, etc., may indeed prove facilitative in tasks requiring deeper processing. Future directions will investigate the influence of NF and NSN on tasks requiring greater depth of processing, such as semantic decision, and test alternative coding schemes for the classification of abstract features.

ALTERNATIVE TASKS

Although semantic richness effects can be detected across a variety of tasks, task-specific differences can reveal important insights into the structure of semantic memory. In studies using concrete stimuli, NF effects have been observed for standard lexical decision, go/no-go lexical decision, progressive demasking, and semantic classification, while semantic neighborhood density only accounts for unique variance in standard and go/no-go lexical decision (Pexman et al., 2008; Yap et al., 2011, 2012). This difference has been attributed to feedback from orthography, as well as to differences in task demands. Because lexical decision requires only a familiarity judgment, the more neighbors the better: every additional neighbor serves as evidence that the target is in fact a word, and the combined activation of many such neighbors speeds the decision (as long as such neighbors are sufficiently distant, cf. Mirman and Magnuson, 2008). If linguistic associates are a core part of the representations of abstract concepts in a way that they are not for concrete ones, NSN may pattern differently for abstract words on deep processing tasks such as semantic classification.

Alternatively, if features represent *ad-hoc* verbal descriptions of the content of simulations and simulations for abstract concepts are activated relatively slowly due to their greater complexity (Barsalou and Wiemer-Hastings, 2005), this could provide an alternate explanation of why NF facilitates lexical decision for concrete but not abstract stimuli. If this is the case, then NF might facilitate semantic classification of abstract concepts, even though it had no measurable effect on abstract concept LDTs. Further study with a greater variety of semantic tasks could shed light on these intriguing questions.

ALTERNATIVE CODES

Any classification scheme that attempts to shoehorn a rich conceptual space into a set of discrete and non-overlapping categories faces significant limitations, and ours is no exception. For example, in the property generation task conducted by Barsalou and Wiemer-Hastings (2005), utterances tagged with code EVC ("any act of communication," p. 159) were far more frequent in participants' descriptions of abstract concepts, and there is some evidence to suggest that abstract concepts may be primarily grounded in acts of communication (Borghi and Cimatti, 2009; Della Rosa et al., 2010; Borghi et al., 2011). Therefore, we wished to include a category (Code 1) encompassing communicative acts (e.g., *explain, demand, call, shout*) and entities with a communicative purpose (e.g., *instructions*, *messages*, *conversation, recommendation, argument*), as this seemed likely to be a type of feature that might participate in abstract concept representations. As many abstract concepts are themselves communicative terms, this category often overlapped with code 19: taxonomic superordinates/subordinates. Due to the taxonomic ambiguity of these terms (is an *inquiry* a kind of *request*?) and the relatively low theoretical relevance of taxonomic relationships to abstract concept representations, such conflicts were resolved by defining code 19 as "hypernyms and hyponyms not otherwise coded." These were seemingly sensible decisions that resulted in high interrater reliability due to ease of coding: All words that described communicative acts or entities with a communicative purpose were tagged with code 1. However, it also meant that the *communication* category became populated with a mishmash of synonyms (yell → **holler**), hypernyms (rule → **decree**), functions (phone → **communicate**), terms that occur when participants describe situations relevant to the concept or its opposite (truth → **lie**), etc. This example alone should make it clear that many possible alternative coding schemes are possible. Although we found no influence of NF on lexical decision for abstract words, alternative methods of subdividing the feature space may reveal feature categories for which a higher NFs does facilitate abstract LDTs.

ALTERNATIVE APPROACHES

While it is possible that we did not detect NF effects for abstract concepts due to the wrong task or the wrong codes, it is also possible that counting features is simply not a useful method for uncovering the structure of abstract concepts. Indeed, our finding that rich linguistic contexts facilitate LDT moreso for abstract words than for highly concrete words is consistent with theoretical claims that language plays a central role in abstract concept representations (Paivio, 1986; Crutch and Warrington, 2005; Borghi et al., 2011), as well as with neuroimaging meta-analyses showing greater activation in language areas during abstract concept processing (Binder et al., 2009; Wang et al., 2010). What might such language-based representations look like? One promising answer comes from corpus-based models of semantic memory such as LSA (Landauer and Dumais, 1997), which construct semantic representations on the basis of distributional statistics. Several computational modelers have demonstrated that superior performance can be achieved on various tasks by extending distributional models with sensorimotor information for concrete concepts (Howell et al., 2005; Jones and Recchia, 2010; Steyvers, 2010; Johns and Jones, 2012); abstract concepts are indirectly grounded in such models by virtue of their linguistic relationships with (grounded) concrete concepts. Alternatively, the corpus-based model of Vigliocco et al. (2009) directly grounds abstract concepts in a combination of linguistic and affective information.

Other approaches include the work of Schmid (2000), who conducted an intensive corpus-based study that elucidates connections between the syntactic and semantic properties of a wide range of abstract nouns and presents an in-depth taxonomy of abstract concept types. Yet others have investigated abstract concepts using such diverse lenses as metaphor (Lakoff, 2009), force dynamics (Talmy, 1988), and many others (see Pecher et al., 2011, for a review). Feature generation should be considered merely one of many possible tools for investigating the nature of abstract concept representations.

CONCLUSION

Questions about the role of context in abstract concept representation go back at least as far as Schwanenflugel and Shoben (1983). Ultimately, the greatest benefit of feature norms for a large set of abstract and concrete concepts may be a better understanding of the precise role that context plays in scaffolding word meaning. The present studies suggest that, at least in lexical decision, NSN facilitates activation of abstract concepts, while NFs facilitates activation of concrete concepts. Analysis of two datasets of feature generation data provided converging evidence that the number of *entity properties* and *concrete situation properties* (i.e., *locations* and *associated entities*) primarily drove our NF effects. A broad interpretation of these results consistent with some theories of concept representation is that while rich language contexts facilitate abstract concept recognition, rich physical characteristics, and contexts facilitate concrete concept recognition. Similar investigations using different tasks are likely to add further nuance to our understanding of different forms of semantic richness and the conceptual representations they support.

ACKNOWLEDGMENTS

This research was supported by NSF BCS-1056744 and a grant from Google Research to Michael N. Jones.

REFERENCES

Adelman, J. S., Brown, G. D. A., and Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychol. Sci.* 17, 814–823.

Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single trial event-related potentials. *Neuropsychologia* 49, 970–983.

Andrews, M., Vigliocco, G., and Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychol. Rev.* 116, 463–498.

Ashcraft, M. H. (1978). Feature dominance and typicality effects in feature statement verification. *J. Verbal Learn. Verbal Behav.* 17, 155–164.

Audet, C., and Burgess, C. (1999). "Using a high-dimensional memory model to evaluate the properties of abstract and concrete words," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, (Vancouver, BC: Canada), 37–42.

Balota, A. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459.

Baroni, M., and Lenci, A. (2008). Concepts and properties in word spaces. *Riv. Linguist.* 20, 55–88.

Barry, C. (1984). "Consistency and semantic errors in a deep dyslexic patient," in *Dyslexia: A Global Issue,* eds R. N. Malatesh and H. J. Whitaker (The Hague: Martinus Nijhoff), 311–337.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2008). "Language and simulation in conceptual processing," in *Symbols, Embodiment, and Meaning,* eds M. De Vega, A. M. Glenberg, and A. C. Graesser (Oxford: Oxford University Press), 245–283.

Barsalou, L. W., and Wiemer-Hastings, K. (2005). "Situating abstract concepts," in *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thought,* eds D. Pecher and R. Zwaan (New York, NY: Cambridge University Press), 129–163.

Binder, J. R., Desai, R., Graves, W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 29, 2767–2796.

Borghi, A. M., and Cimatti, F. (2009). "Words as tools and the problem of abstract word meanings," in *Proceedings of the Annual Conference of the Cognitive Science Society,* eds N. Taatgen and H. van Rijn, (Amsterdam), 2304–2309.

Borghi, A. M., Flumini, A., Cimatti, F., Marocco, D., and Scorolli, C. (2011). Manipulating objects and telling words: a study on concrete and abstract words acquisition. *Front. Psychology* 2:15. doi: 10.3389/fpsyg.2011.00015

Brysbaert, M., and New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990.

Buchanan, L., Westbury, C., and Burgess, C. (2001). Characterizing semantic space: neighborhood effects in word recognition. *Psychon. Bull. Rev.* 8, 531–544.

Bullinaria, J. A., and Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: a computational study. *Behav. Res. Methods* 39, 510–526.

Coltheart, M. (1981). The MRC psycholinguistic database. *Q. J. Exp. Psychol.* 33, 497–505.

Crutch, S. J., and Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain* 128, 615–627.

de Mornay Davies, P., and Funnell, E. (2000). Semantic representation and ease of predication. *Brain Lang.* 73, 92–119.

Della Rosa, P. A., Catricalà, E., Vigliocco, G., and Cappa, S. F. (2010). Beyond the abstract–concrete dichotomy: mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behav. Res. Methods* 42, 1042–1048.

Duñabeitia, J. A., Avilés, A., and Carreiras, M. (2008). NoA's ark: influence of the number of associates in visual word recognition. *Psychon. Bull. Rev.* 15, 1072–1077.

Durda, K., Buchanan, L., and Caron, R. (2006). *WordMine2 [Online].* Available online at: www.wordmine2.org

Gilhooly, K. J., and Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behav. Res. Methods Instrum.* 12, 395–427.

Grondin, R., Lupker, S. J., and McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *J. Mem. Lang.* 60, 1–19.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *J. Verbal Learn. Verbal Behav.* 18, 441–461.

Hare, M., Jones, M., Thomson, C., Kelly, S., and McRae, K. (2009). Activating event knowledge. *Cognition* 111, 151–167.

Howell, S. R., Jankowicz, D., and Becker, S. (2005). A model of grounded language acquisition: sensorimotor features improve grammar learning. *J. Mem. Lang.* 53, 258–276.

Johns, B. T., and Jones, M. N. (2012). Perceptual inference from global lexical similarity. *Top. Cogn. Sci.* 4, 103–120.

Jones, G. V. (1985). Deep dyslexia, imageability and ease of predication. *Brain Lang.* 24, 1–19.

Jones, M. N., Johns, B. T., and Recchia, G. L. (2012). The role of semantic diversity in lexical organization. *Can. J. Exp. Psychol.* 66, 121–132.

Jones, M. N., and Recchia, G. L. (2010). "You can't wear a coat rack: a binding framework to avoid illusory feature migrations in perceptually grounded semantic models," in *Proceedings of the 32nd Annual Cognitive Science Society,* eds S. Ohlsson and R. Catrambone (Portland, OR), 877–882.

Kiefer, M., and Pulvermüller, F. (2011). Conceptual representations in mind and brain: theoretical developments, current evidence, and future directions. *Cortex* 48, 805–825.

Kievit-Kylar, B., and Jones, M. N. (2011). "The semantic pictionary project," in *Proceedings of the 33rd Annual Conference of the Cognitive Science Society,* eds L. Carlson, C. Hölscher, and T. Shipley (Boston, MA), 2229–2234.

Kiss, G. R., Armstrong, C., Milroy, R., and Piper, J. (1973). "An associative thesaurus of English and its computer analysis," in *The Computer and Literary Studies,* eds A. J. Aitkin, R. W. Bailey, and N. Hamilton-Smith (Edinburgh: Edinburgh University Press), 153–165.

Lakoff, G. (2009). "The neural theory of metaphor," in *The Cambridge Handbook of Metaphor and Thought* (Cambridge: Cambridge University).

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychol. Rev.* 104, 211–240.

Louwerse, M. (2008). Embodied relations are encoded in language. *Psychon. Bull. Rev.* 15, 838–844.

Lund, K., and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* 28, 203–208.

Manning, C. D., and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing,* Cambridge MA: MIT Press.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behav. Res. Methods* 37, 547–559.

Medin, D. L. (1989). Concepts and conceptual structure. *Am. Psychol.* 44, 1269–1481.

Miller, G. A. (1990). WordNet: an online lexical database. *Int. J. Lexicogr.* 3, 235–312.

Mirman, D., and Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: processing is slowed by near neighbors and speeded by distant neighbors. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 65–79.

Moss, H. E., and Tyler, L. K. (1995). Investigating semantic memory impairments: the contribution of semantic priming. *Memory* 3, 359–395.

Nelson, D. L., McEvoy, C. L., and Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Available online at: http://w3.usf.edu/FreeAssociation/

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford, England: Oxford University Press.

Paivio, A. (2010). Dual coding theory and the mental lexicon. *Ment. Lexicon* 5, 205–230.

Pecher, D., Boot, I., and van Dantzig, S. (2011). "Abstract concepts: sensory-motor grounding, metaphors, and beyond," in *The Psychology of Learning and Motivation, Vol. 54,* ed B. Ross (Burlington: Academic Press), 217–248.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., and Pope, J. (2008). There are many ways to be rich: effects of three measures of semantic richness on visual word recognition. *Psychon. Bull. Rev.* 15, 161–167.

Pexman, P. M., Holyk, G. G., and Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Mem. Cognit.* 31, 842–855.

Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number-of-features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.

Plaut, D. C., and Shallice, T. (1993). Deep dyslexia: a case-study of connectionist neuropsychology. *Cogn. Neuropsychol.* 10, 377–500.

Rabovsky, M., Sommer, W., and Rahman, R. A. (2012). The time course of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:11. doi: 10.3389/fnhum.2012.00011

Recchia, G., and Jones, M. N. (2009). More data trumps smarter algorithms: training computational models of semantics on very large corpora. *Behav. Res. Methods* 41, 647–656.

Recchia, G., and Jones, M. N. (2012). Supplementary materials [Online]. Available online at: http://www.indiana.edu/∼clcl/RJ_supplementary_materials.pdf

Recchia, G., Kievit-Kylar, B., Jones, M. N., and McRae, K. (2011). "Using web games to elicit associative and feature-based conceptual representations," in *Annual Meeting of the Society for Computers in Psychology*. (Seattle, WA).

Riordan, B., and Jones, M. N. (2011). Redundancy in linguistic and perceptual experience: comparing distributional and feature-based models of semantic representation. *Top. Cogn. Sci.* 3, 303–345.

Rosch, E., and Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cogn. Psychol.* 7, 573–605.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., and Binder, J. R. (2005). Modulation of the semantic system by word imageability. *Neuroimage* 27, 188–200.

Saffran, E. M. (1980). Reading in deep dyslexia is not ideographic. *Neuropsychologia* 18, 219–223.

Schmid, H.-J. (2000). *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition.* New York, NY: Mouton de Gruyter.

Schwanenflugel, P. J., and Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *J. Exp. Psychol. Learn. Mem. Cogn.* 9, 82–102.

Shaoul, C., and Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behav. Res. Methods* 42, 393–413.

Sigman, M., and Cecchi, G. A. (2002). Global organization of the Wordnet lexicon. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1742–1747.

Singh, P., Lin, T., Mueller, E. T., Lim, G., Perkins, T., and Zhu, W. L. (2002). "Open mind commonsense: knowledge acquisition from the general public," in *Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems, Lecture Notes in Computer Science No. 2519*, (Heidelberg: Springer).

Speer, R., Havasi, C., and Surana, H. (2010). "Using verbosity: common sense data from games with a purpose," in *Proceedings of the Florida Artificial Intelligence Research Society (FLAIRS) Conference,* (Florida: Daytona Beach).

Steyvers, M. (2010). Combining feature norms and text data with topic models. *Acta Psychol.* 133, 234–342.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognit. Sci.* 12, 49–100.

Tyler, L. K., Moss, H. E., Galpin, A., and Voice, J. K. (2002). Activating meaning in time: the role of imageability and form-class. *Lang. Cogn. Process.* 17, 471–502.

Vigliocco, G., Meteyard, L., Andrews, M., and Kousta, S. (2009). Toward a theory of semantic representation. *Lang. Cogn.* 1, 219–248.

Vinson, D. P., and Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behav. Res. Methods* 40, 183–190.

von Ahn, L. (2006). Games with a purpose. *Computer* 39, 92–94.

Wang, J., Conder, J. A., Blitzer, D. N., and Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: a meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 31, 1459–1468.

Wiemer-Hastings, K., and Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognit. Sci.* 29, 719–736.

Wu, L. L., and Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychol.* 132, 173–189.

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., and Huff, M. J. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* 6:72. doi: 10.3389/fnhum.2012.00072

Yap, M. J., Tan, S. E., Pexman, P. M., and Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychon. Bull. Rev.* 18, 742–750.

Yarkoni, T., Balota, D. A., and Yap, M. J. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychon. Bull. Rev.* 15, 971–979.

Zeno, S., Ivens, S., Millard, R., and Duvvuri, R. (1995). *The Educator's Word Frequency Guide*. Brewster, NJ: Touchstone Applied Science Associates, Inc.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 September 2012; accepted: 07 November 2012; published online: 27 November 2012.

Citation: Recchia G and Jones MN (2012) The semantic richness of abstract concepts. Front. Hum. Neurosci. ***6****:315. doi: 10.3389/fnhum.2012.00315*

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 08 October 2012
doi: 10.3389/fnhum.2012.00275

Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns

P. Ian Newcombe[1], Cale Campbell[1], Paul D. Siakaluk[1] and Penny M. Pexman[2]*

[1] University of Northern British Columbia, Prince George, BC, Canada
[2] University of Calgary, Calgary, AB, Canada

Edited by:
Melvin Yap, National University of Singapore, Singapore

Reviewed by:
Gabriella Vigliocco, University College London, UK
David Vinson, University College London, UK
Christy Wilson-Mendenhall, Northeastern University, USA

***Correspondence:**
Paul D. Siakaluk, Department of Psychology, 3333 University Way, University of Northern British Columbia, Prince George, BC, Canada, V2N 4Z9.
e-mail: siakaluk@unbc.ca

There is much empirical evidence that words' relative imageability and body-object interaction (BOI) facilitate lexical processing for concrete nouns (e.g., Bennett et al., 2011). These findings are consistent with a grounded cognition framework (e.g., Barsalou, 2008), in which sensorimotor knowledge is integral to lexical processing. In the present study, we examined whether lexical processing is also sensitive to the dimension of emotional experience (i.e., the ease with which words evoke emotional experience), which is also derived from a grounded cognition framework. We examined the effects of emotional experience, imageability, and BOI in semantic categorization for concrete and abstract nouns. Our results indicate that for concrete nouns, emotional experience was associated with less accurate categorization, whereas imageability and BOI were associated with faster and more accurate categorization. For abstract nouns, emotional experience was associated with faster and more accurate categorization, whereas BOI was associated with slower and less accurate categorization. This pattern of results was observed even with many other lexical and semantic dimensions statistically controlled. These findings are consistent with Vigliocco et al.'s (2009) theory of semantic representation, which states that emotional knowledge underlies meanings for abstract concepts, whereas sensorimotor knowledge underlies meanings for concrete concepts.

Keywords: emotional experience, imageability, body-object interaction, semantic richness, grounded cognition

INTRODUCTION

Classical theories of cognition hold that perception and cognition have distinct representational formats, such that perceptual representations are modal, whereas conceptual representations are amodal (Fodor, 1983; Pylyshyn, 1984). More recently, a growing number of cognitive scientists have proposed an alternative theoretical perspective, called embodied or grounded cognition, in which modal representations underlie both perceptual and conceptual knowledge (e.g., Pecher and Zwaan, 2005). One well-developed and influential grounded cognition framework is perceptual symbol systems (PSS: Barsalou, 1999). According to PSS, conceptual knowledge is largely acquired through bodily interaction with the environment and is inherently multimodal, such that different aspects of conceptual knowledge are stored in different neural systems dedicated to sensorimotor processing (e.g., sensory knowledge is stored in neural systems dedicated to sensory processing, whereas motor knowledge is stored in neural systems dedicated to motor processing). Conceptual processing occurs via simulation, or the partial reenactment of the various neural states that were involved during bodily interaction with the environment. More recently, Barsalou (2003, 2008, 2009) has emphasized that an important aspect of the acquisition and subsequent simulation of conceptual knowledge is that it does not occur in a contextual vacuum. That is, "(a)t any given moment in perception, people perceive the immediate space around them, including agents, objects, and events present" (Barsalou, 2009, p. 1283). Barsalou (2003) referred to the fact that conceptual knowledge is influenced by environmental context as *situated conceptualization.*

The PSS framework has recently been extended in empirical efforts investigating whether sensorimotor knowledge influences lexical processing. Several studies have examined the influence of knowledge gained through sensory experience, as measured by imageability (i.e., how easily words evoke mental images), in lexical processing. Imageability has been shown to facilitate responding in lexical decision, word naming, picture naming, progressive demasking, and semantic categorization tasks (Balota et al., 2004; Bennett et al., 2011; Yap et al., 2012). Additional studies have examined the influence of knowledge gained through motor experience in lexical processing, as indexed by a dimension known as body-object interaction (BOI), which measures perceptions of the ease with which a human body can physically interact with a word's referent. BOI has been shown to facilitate responding in lexical decision, phonological lexical decision, word naming, picture naming, semantic categorization, and sentence processing tasks (Siakaluk et al., 2008a,b; Tillotson et al., 2008; Bennett et al., 2011; Wellsby et al., 2011; Hansen et al., 2012; Phillips et al., 2012; Tousignant and Pexman, 2012; Yap et al., 2012). Further, in an fMRI study involving the semantic categorization task (SCT), processing of high BOI words was associated with greater activation in the left inferior parietal lobule (supramarginal gyrus, BA 40), a sensory association area involved in kinesthetic memory (Hargreaves et al., 2012). This finding suggests that knowledge about the relative availability of motor experience with words' referents is activated during visual word recognition. The facilitatory effects of imageability and BOI arise, according to the PSS framework, because easily imageable words and high BOI words refer to concrete concepts that occur in environmental contexts that allow development of relatively rich stores of sensory knowledge and motor knowledge, and elicit richer sensory simulations and motor simulations. That is, by virtue of their associated sensory knowledge and motor knowledge, easily imageable words and high BOI words enjoy *semantic richness*.

The semantic feedback activation framework describes in more explicit detail how semantic richness effects, such as the facilitatory effects of imageability and BOI, arise within the visual word recognition system (Hino and Lupker, 1996; Pexman et al., 2002). According to this framework, the visual word recognition system is comprised of separate but interconnected sets of units dedicated to processing orthographic, phonological, and semantic information. In tasks in which responses are based primarily on orthographic processing (e.g., lexical decision), semantically richer words (i.e., easily imageable words and high BOI words) generate greater levels of semantic activation (i.e., richer sensory simulations and richer motor simulations) within semantic units, which leads to greater semantic feedback activation to orthographic units, thus leading to faster responding for these words. The same explanation holds in tasks in which responses are based primarily on phonological processing (e.g., word naming), except that the relevant semantic feedback activation is that which influences phonological units. In tasks in which responses are based primarily on semantic processing (e.g., semantic categorization), greater levels of semantic activation (i.e., richer sensory simulations and richer motor simulations) generated by semantically richer words (i.e., easily imageable words and high BOI words) leads to faster settling of their associated semantic representations within semantic units. Notably, while many studies have reported facilitatory effects of various semantic richness dimensions in SCTs, it has also been established that the nature of these effects will depend on the particular decision category involved, and how the richness dimension is relevant to that category (e.g., Pexman et al., 2003; Tousignant and Pexman, 2012).

GROUNDING LEXICAL SEMANTICS BEYOND SENSORIMOTOR KNOWLEDGE

Recently, it has been noted that researchers with a grounded cognition perspective have focused primarily on the sensorimotor aspects of cognition, and have largely ignored other types of knowledge that are relevant to this perspective, such as the emotional aspects of cognition (Vigliocco et al., 2009; Parisi, 2011). This is an important consideration, because there is accumulating evidence that emotional knowledge plays a number of roles in cognition more generally (Dolan, 2002; Vigliocco et al., 2009), and in conceptual processing more particularly (Niedenthal et al., 2005a,b, 2009; Wilson-Mendenhall et al., 2011).

Wilson-Mendenhall et al. (2011) extended the PSS framework to account for how emotional concepts may be acquired through bodily experience via situated conceptualization. That is, they suggested that emotional knowledge is acquired and simulated in the same manner in which sensorimotor knowledge is acquired and simulated. They stated, "Like all concepts, emotion concepts originate and operate in the context of continuous situated activity, with situations typically including a physical setting, agents, objects, and actions in the world, interoceptive sensations from the body, and mentalizing related to prospective and retrospective thought" (p. 1108). Thus, although emotional knowledge can be considered more abstract in nature than sensorimotor knowledge, both types of knowledge are nonetheless acquired and simulated through the same mechanisms. More specifically, emotional knowledge is stored in the neural systems dedicated to emotional processing, and conceptual processing of emotional knowledge occurs through simulation, or the partial reenactment of the various neural states that were involved during bodily interaction with the environment.

Vigliocco et al. (2009) developed a similar framework, and argued that emotional knowledge is grounded in bodily experience and is integral to conceptual processing. According to Vigliocco et al.'s (2009) framework of semantic representation, there are two general classes of knowledge that humans are able to acquire and use in conceptual processing. One type is what they refer to as *experiential* knowledge and the other type is what they refer to as *linguistic* knowledge. They characterize experiential knowledge as being derived not only from sensory and motor experience with the external environment, but also from affective or emotional experience with the internal environment of one's own body (e.g., the experience of moods or feelings) in conjunction with external events. They characterize linguistic knowledge as lexical co-occurrence information (e.g., the Topics model, Griffiths et al., 2007; the LSA model, Landauer and Dumais, 1997; the HAL model, Lund and Burgess, 1996), and syntactic information[1]. Importantly, they further propose that sensorimotor knowledge is more salient to the acquisition and use of concrete concepts than abstract concepts, whereas emotional knowledge and linguistic knowledge types are more salient to the acquisition and use of abstract concepts than concrete concepts. In other words, sensorimotor knowledge can be thought of as diagnostic of concrete concepts, whereas emotional knowledge and linguistic knowledge can be thought of as diagnostic of abstract concepts.

In support of the idea that emotional knowledge is integral to the processing of abstract concepts in lexical processing, Kousta et al. (2009) reported faster lexical decision latencies for negative words and for positive words than for neutral words (there was no difference in latencies between the negative words and the positive words). Importantly, these facilitatory effects of emotion were independent of any influence of concreteness or imageability (all three sets of words were matched on these two dimensions). In addition, Kousta et al. (2011) reported an intriguing result they called the *abstractness effect*. When imageability and context availability were statistically controlled, Kousta et al. (2011) reported that lexical decision latencies were faster for abstract words than for concrete words, in contrast to the typical finding whereby latencies are faster for concrete words than for abstract words. They attributed this reversal to the facilitatory influence of emotional

[1] See Barsalou et al. (2008) for a related framework they call Language and Situated Simulation (LASS).

content for abstract words. Thus, both Vigliocco et al.'s (2009) and Wilson-Mendenhall et al.'s (2011) frameworks conceive emotional knowledge as situationally and experientially derived, and we use those characterizations to frame predictions in the current study.

THE PRESENT STUDY

The primary purpose of the present study was to examine the relative contributions of emotional, sensory, and motor knowledge to semantic processing for concrete and abstract nouns. In doing so, our objective was to test Vigliocco et al.'s (2009) claims that emotional knowledge underlies the meanings of abstract nouns, whereas sensory and motor knowledge underlies the meanings of concrete nouns. More specifically, we assessed emotional knowledge using a new dimension we call *emotional experience*, which was designed to capture the relative ease with which words elicit or evoke emotional experience. We assessed sensory knowledge using the dimension of imageability, and we assessed motor knowledge using the dimension of BOI. We examined the behavioral effects of the above three dimensions in semantic categorization for nouns referring to concrete concepts and for nouns referring to abstract concepts. We characterized emotional experience as a unitary dimension to make it analogous to imageability and BOI, and thus to facilitate comparisons between these three dimensions of experiential knowledge.

An example may help elucidate how different types of experiential knowledge may underlie the development and activation of concrete and abstract conceptual knowledge. Imagine a situation in which you are very thirsty and have been looking for some time for something to drink. When you finally see a fountain, you run to it, take a good drink of water, and feel relieved. This situation involves the concrete concept of "fountain" and the abstract concept of "relief." The concrete noun *fountain* can be considered easily imageable, because it refers to things that can be easily perceived by the senses (e.g., sight, touch), and high on the BOI dimension, because it also refers to things that can be easily physically interacted with (e.g., turning the knob to start the flow of water, holding the fountain for balance, bending down to drink the water). We suggest that *fountain* can be considered low on the emotional experience dimension, because it is unlikely that it refers to things that are reliably associated with relatively robust emotional experiences (e.g., when one sees fountains, are they always associated with emotional experiences as they are with visual experiences and motor experiences?). Thus, what people know about the concrete concept "fountain" will be derived primarily from sensory experience and motor experience, and perhaps only secondarily from emotional experience. Conversely, the abstract noun *relief* can be considered high on the emotional experience dimension, because it refers to the alleviation or deliverance from distress (e.g., drinking water to quench thirst). *Relief* can also be considered not easily imageable, because it cannot be easily perceived by the senses, and low on the BOI dimension, because it cannot be easily physically interacted with (if at all). Thus, what people know about the abstract concept "relief" will be derived primarily from emotional experience, and perhaps only secondarily (if at all) from sensory experience and motor experience.

According to Vigliocco et al.'s (2009) framework of semantic representation, there is a "statistical preponderance for sensory-motor information to underlie concrete word meanings and a preponderance for affective (i.e., emotional)...information to underlie abstract word meanings" (p. 223). The implications of these claims are that sensory knowledge and motor knowledge should be especially salient to the processing of concrete nouns, whereas emotional knowledge should be especially salient to the processing of abstract nouns. We selected our stimuli to be either concrete or abstract (described below), and presented these stimuli in two separate tasks. In the concrete SCT, the decision criterion was to decide if the stimuli referred to concrete nouns, whereas in the abstract SCT, the decision criterion was to decide if the stimuli referred to abstract nouns. We propose that there are two benefits in using this experimental design. First, the decision criteria allowed for a relatively more pure assessment of the effects of the three dimensions of experiential knowledge than would a task such as lexical decision. This is because each SCT directly emphasizes processing of the relevant semantic characteristic under examination, namely concreteness in the concrete SCT and abstractness in the abstract SCT, rather than requiring a more peripheral decision to be made in lexical decision (e.g., is the item a word). Second, the decision criteria allowed a more refined analysis of the individual effects of each of the three dimensions of experiential knowledge in each SCT. That is, according to the semantic feedback activation framework, the nature of the effects of any given semantic richness dimension in the SCT will depend on whether the dimension is congruent with the decision category (Pexman et al., 2003; Tousignant and Pexman, 2012).

As such, we made the following prediction regarding the effects of imageability and BOI on the processing of concrete nouns in the concrete SCT: both dimensions should facilitate categorization, such that higher ratings on these two dimensions should be associated with faster and more accurate categorizations, because these two dimensions are diagnostic of concrete concepts, and thus are congruent with the decision criterion of "is the word concrete?". We made the following two more speculative predictions regarding the effects of emotional experience on the processing of concrete concepts in the concrete SCT. One possibility, derived from the semantic feedback activation framework of visual word recognition (Hino and Lupker, 1996; Pexman et al., 2002), is that because emotional experience is diagnostic of abstract concepts, it thus may inhibit categorization in the concrete SCT, because this dimension is not congruent with the decision criterion of "is the word concrete?". An alternative possibility is that no effects of this dimension will be observed, and this may arise either because the effects of emotional experience may be too subtle to detect using the type of experimental design employed in the present study (although they may be detectable using other experimental tasks), or they play little or no role in the processing of concrete concepts (admittedly, this is a very strong interpretation of Vigliocco et al.'s (2009), framework of semantic representation), or for some other reason.

We made the following prediction regarding the effects of emotional experience on the processing of abstract nouns in the abstract SCT: emotional experience should facilitate categorization, such that higher ratings on this dimension should be associated with faster and more accurate categorizations, because emotional experience is diagnostic of abstract concepts, and thus is congruent with the decision criterion of "is the word abstract?". We made the following two more speculative predictions regarding the effects of imageability and BOI on the processing of abstract concepts in the abstract SCT. One possibility, derived from the semantic feedback activation framework of visual word recognition (Hino and Lupker, 1996; Pexman et al., 2002), is that because these two dimensions are diagnostic of concrete concepts, they thus may inhibit categorization in the abstract SCT, because these dimensions are not congruent with the decision criterion of "is the word abstract?". An alternative possibility is that no effects of these two dimensions will be observed, and this may arise for the same reasons outlined in the paragraph above (i.e., the effects of imageability and BOI may be too subtle to detect using the type of experimental design employed in the present study, or they play little or no role in the processing of abstract concepts, or for some other reason).

MATERIALS AND METHODS

PARTICIPANTS

Two separate groups of 30 undergraduate students from the University of Northern British Columbia participated for bonus course credit: one group participated in the concrete SCT and the other group participated in the abstract SCT. All were native English speakers and reported normal or corrected-to-normal vision.

STIMULI

Two hundred concrete nouns and 200 abstract nouns were selected from the Toronto Word Pool (Friendly et al., 1982) or the Paivio et al. (1968) word banks. The concrete and abstract nouns are listed in Concrete Nouns Used in the Experiments and Abstract Nouns Used in the Experiments in Appendix, respectively. Concrete nouns had concreteness and imageability ratings of 5.0 or higher, whereas abstract nouns had concreteness and imageability ratings of 3.9 or less. The concrete nouns and the abstract nouns were matched pairwise on print length. Values were obtained for the following control variables: HAL log-frequency, Levenshtein orthographic distance, number of letters, phonemes, syllables, and morphemes (all taken from Balota et al., 2007), age of acquisition (AoA, taken from Kuperman et al., 2012), concreteness[2], to control for typicality effects (as noted, taken from either Friendly et al. or Paivio et al.), number of senses (retrieved from the www.wordsmyth.net), and the inverse of the number of neighbor words (plus 1) within the neighborhood threshold (NCOUNT-INV)[3] (Shaoul and Westbury, 2010a,b). The semantic richness variables of interest for the present study included: emotional experience, imageability, and BOI. Emotional experience ratings and BOI ratings were collected from two separate groups of 30 undergraduate students from the University of Calgary. The instructions used for the emotional experience ratings were derived for the present study and are given in Written Instructions Used for the Emotional Experience Rating Task in Appendix. The instructions used for the BOI ratings were the same as those used by Tillotson et al. (2008).

APPARATUS AND PROCEDURE

The 200 concrete nouns and the 200 abstract nouns were presented in both tasks. For the concrete SCT, participants were instructed to decide only whether each word referred to a concrete noun, and to respond by pressing the "?" key on the computer keyboard if the word did refer to a concrete noun and to not press any key if the word did not refer to a concrete noun (i.e., participants were instructed to respond only to the concrete nouns). For the abstract SCT, participants were instructed to decide only whether each word referred to an abstract noun, and to respond by pressing the "?" key if the word did refer to an abstract noun and to not press any key if the word did not refer to an abstract noun (i.e., participants were instructed to respond only to the abstract nouns). Participants were instructed to make their responses as quickly and as accurately as possible, and were told that the stimuli for which they did not make a response would be automatically replaced by the next stimulus item after 2500 ms. The stimuli were presented in the center of a color VGA monitor driven by a Pentium-class microcomputer running DirectRT software[4]. A trial was initiated by a fixation marker that appeared at the center of the computer display for 1000 ms and was then replaced by a stimulus item. The intertrial interval was 2000 ms. Stimulus order was randomized separately for each participant. Following every 100 trials, participants had an opportunity to take a break, and continued when ready by pressing the spacebar. Before beginning either task,

[2]Several studies have demonstrated the importance of typicality in semantic categorization. For example, Hampton (1997) reported that typicality accounted for significant amounts of unique within-category variability in both categorization latency and response probability across a variety of categories (see also, Casey, 1992; Larochelle and Pineau, 1994; and Smith et al., 1974, for similar results). Because the two categories used in the present study were quite broad, namely whether nouns were concrete or abstract, we used the concreteness ratings from the Friendly et al. (1982) and the Paivio et al. (1968) norms as typicality ratings. The instructions used in these studies included the following, "Each word was accompanied by a seven-point bipolar numerical scale, with the extremes labeled Highly Abstract and Highly Concrete, respectively...the ends of the scale were defined in terms of abstractness-concreteness rather than low concreteness-high concreteness...the present instructions stated that, 'Any word that refers to objects, materials, or persons should receive a *high concreteness* rating; any word that refers to an abstract concept that cannot be experienced by the senses should receive a *high abstractness* rating"' (Paivio et al., 1968, p. 5). Thus, the lower ratings associated with the abstract words provided an indication of their relative "abstractness" (as was indicated at the lower end of the scale), and are therefore appropriate for entry into the analysis of the abstract categorization task data. The higher ratings associated with the concrete words provided an indication of their relative "concreteness" (as was indicated at the higher end of the scale), and are therefore appropriate for entry into the analysis of the concrete categorization task data. Inclusion of these ratings is important, because they will account for within-category variability, and therefore allow for a more stringent test of whether the three dimensions of experiential knowledge account for additional categorization latency and error variability, above and beyond that accounted for by the concreteness ratings (and the other variables entered in the first step of the analyses).

[3]Shaoul and Westbury (2010a) developed the High Dimensional Explorer (HiDEx) model of lexical co-occurrence, based on the basic architecture of the HAL model (Lund and Burgess, 1996). One of the measures that HiDEx computes is called NCOUNT-INV. This measure is the inverse (plus 1) of another measure HiDEx computes, called NCOUNT, which is the number of neighbor words within a specified neighborhood membership threshold. NCOUNT-INV "has a value of 1 for words with no neighbors and smaller values for words with more neighbors" (p. 397). It is essentially a measure of the lexical co-occurrence neighborhood size for a given word. We used the NCOUNT-INV measure because Shaoul and Westbury (2010a) reported that it best correlated with the SCT they used in their study (the decision category was whether words referred to living things).

[4]http://www.empirisoft.com/DirectRT.aspx

participants were given practice trials consisting of 10 concrete nouns and 10 abstract nouns.

DATA ANALYSIS

The data from both tasks were first analyzed jointly to test for interaction effects between task (i.e., concrete, abstract) and each of emotional experience, imageability, and BOI. The following variables were entered in the first step of a hierarchical multiple regression analysis: HAL log-frequency, AoA, Levenshtein orthographic distance, number of phonemes, syllables, and morphemes, concreteness, number of senses, NCOUNT-INV, type of task (dummy coded; "1" for concrete nouns, "2" for abstract nouns), and emotional experience, imageability, and BOI[5]. The final three variables were centered prior to inclusion in the analysis (Keith, 2006). The following interaction variables were entered in the second step: task by emotional experience, task by imageability, and task by BOI. These three interaction terms were constructed by creating cross-product terms through the multiplication of the task variable with the appropriate centered semantic richness variable (Keith, 2006). To follow up any significant interactions, we then examined the effects of emotional experience, imageability, and BOI in each data set separately in two additional hierarchical multiple regression analyses. In these follow up analyses, the following variables were entered in the first step: HAL log-frequency, AoA, Levenshtein orthographic distance, number of phonemes, syllables, and morphemes, concreteness, number of senses, and NCOUNT-INV. Emotional experience, imageability, and BOI were then entered in the second step. We used hierarchical multiple regression because it provided two important pieces of information, namely, the change in R^2 when the three dimensions of experiential knowledge were added to the analyses (after a number of control variables were already entered), and whether each of the three dimensions of experiential knowledge accounted for a significant amount of unique variability in semantic processing.

RESULTS

There were 10 concrete nouns (from the concrete SCT data) and 10 abstract nouns (from the abstract SCT data) that had error rates greater than 30%. In addition, the abstract nouns *justice* and *moment* were used as examples in the emotional experience ratings instructions. Therefore, in the omnibus categorization latency and categorization error analyses, the 10 concrete nouns and the 10 abstract nouns with high error rates, along with the two abstract nouns used in the emotional experience ratings instructions, were removed. Further, in the follow up analyses of the concrete SCT data, only the 10 concrete nouns had to be removed, whereas in the follow up analyses of the abstract SCT data, only the 10 abstract nouns and the two abstract nouns used in the emotional experience ratings instructions had to be removed. The items that were removed from the analyses are indicated with * in Concrete Nouns Used in the Experiments and Abstract Nouns Used in the Experiments in Appendix. Outliers were identified in the following manner. First, categorization latencies faster than 250 ms or slower than 2000 ms were considered outliers. Second, for each participant, categorization latencies greater than 2.5 SDs from the mean were considered outliers. Using this procedure, a total of 151 observations (2.6% of the data) were removed from the concrete SCT data set, and a total of 178 observations (3.2% of the data) were removed from the abstract SCT data set. The raw categorization latencies were z score transformed before analysis.

OMNIBUS ANALYSIS

Means and SDs for the predictor variables for the concrete nouns and the abstract nouns are shown in **Table 1** (note that we included the uncentered means for the emotional experience, imageability, and BOI variables, and that the SDs are identical whether using the uncentered or centered means). Zero-order correlations between the criterion variables and the predictor variables for the concrete SCT are presented in **Table 2**, and zero-order correlations between the criterion variables and the predictor variables for the abstract SCT are presented in **Table 3**. For the regression analyses, the critical results are those for the three interaction tests at step 2, and thus only these results are shown in **Table 4**. For both criterion variables (categorization latencies, categorization errors), there was a significant change in R^2 when the three interaction terms were added to the analyses, and importantly, each of the three interaction tests were significant. Interestingly, and consistent with Vigliocco et al.'s (2009) framework of semantic representation, the effects of the imageability and BOI dimensions were in the same direction, whereas the effects of the emotional

Table 1 | Descriptive statistics and behavioral data for the 190 concrete nouns (from the concrete SCT) and the 188 abstract nouns (from the abstract SCT).

Variable	Concrete nouns		Abstract nouns	
	M	**SD**	***M***	**SD**
Log-frequency (HAL)	8.53	1.79	8.88	1.82
Age of acquisition	6.70	1.98	9.65	2.31
Levenshtein orthographic distance	2.60	0.92	2.51	0.64
Letters	7.14	1.75	7.22	1.73
Phonemes	5.67	1.66	6.15	1.65
Syllables	2.24	0.65	2.47	0.84
Morphemes	1.39	0.61	1.70	0.68
Concreteness	6.16	0.52	2.57	0.69
Senses	2.62	1.59	3.21	1.63
NCOUNT-INV	0.23	0.41	0.22	0.41
Emotional experience	2.18	0.77	3.39	1.10
Imageability	5.80	0.63	3.00	0.57
Body-object interaction	4.89	0.93	2.00	0.33
Raw categorization latencies	783.96	115.30	918.32	93.64
Categorization errors	4.09	6.91	5.44	5.94

NCOUNT-INV, inverse of number of word neighbors plus 1.

[5]We did not include number of letters in any of the multiple regression analyses, because of the high zero-order correlations between this variable and the Levenshtein orthographic distance and number of phonemes variables for both the concrete and the abstract noun sets.

Table 2 | Zero-order correlations between the criterion variables and the predictor variables for the concrete SCT.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. CL	–														
2. Errors	0.63**	–													
3. Freq	−0.09	0.04	–												
4. AoA	0.59**	0.32**	−0.31**	–											
5. LOD	0.06	−0.10	−0.63**	0.30**	–										
6. Letters	0.13	−0.01	−0.58**	0.27**	0.89**	–									
7. Phon	0.15*	−0.01	−0.50**	0.29**	0.79**	0.83**	–								
8. Syll	0.10	0.03	−0.42**	0.19**	0.70**	0.72**	0.78**	–							
9. Morph	0.04	−0.06	−0.37**	0.07	0.38**	0.48**	0.39**	0.30**	–						
10. Conc	−0.58**	−0.50**	0.14	−0.43**	−0.35**	−0.39**	−0.46**	−0.37**	−0.24**	–					
11. Senses	0.10	0.13	0.48**	−0.16*	−0.33**	−0.32**	−0.34**	−0.29**	−0.20**	−0.02	–				
12. INV + 1	−0.05	−0.11	−0.70**	0.10	0.59**	0.55**	0.48**	0.38**	0.35**	−0.17*	−0.36**	–			
13. EE	0.04	0.20**	0.39**	−0.09	−0.19**	−0.20**	−0.15*	−0.15*	−0.20**	−0.09	0.17*	−0.30**	–		
14. Image	−0.61**	−0.49**	0.15*	−0.61**	−0.19**	−0.22**	−0.24**	−0.18*	−0.08	0.66**	0.04	−0.08	0.05	–	
15. BOI	−0.62**	−0.64**	−0.02	−0.37**	0.07	0.02	0.00	0.04	0.06	0.49**	−0.12	0.14	0.00	0.41**	–

**p* < 0.05, ***p* < 0.01.

CL, categorization latency; Freq, HAL log-frequency; AoA, age of acquisition; LOD, Levenshtein orthographic distance; Letters, number of letters; Phon, number of phonemes; Syll, number of syllables; Morph, number of morphemes; Conc, concreteness; Senses, number of senses; INV+ 1, inverse of number of word neighbors plus 1; EE, emotional experience; Image, imageability; BOI, body-object interaction.

Table 3 | Zero-order correlations between the criterion variables and the predictor variables for the abstract SCT.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. CL	–														
2. Errors	0.42**	–													
3. Freq	−0.26**	0.02	–												
4. AoA	0.24**	0.00	−0.63**	–											
5. LOD	0.12	−0.07	−0.39**	0.37**	–										
6. Letters	0.12	−0.11	−0.39**	0.39**	0.79**	–									
7. Phon	0.25**	−0.01	−0.42**	0.47**	0.71**	0.84**	–								
8. Syll	0.22**	−0.04	−0.41**	0.49**	0.63**	0.70**	0.75**	–							
9. Morph	0.11	0.01	−0.32**	0.36**	0.49**	0.69**	0.62**	0.64**	–						
10. Conc	0.24**	0.39**	0.18*	−0.12	−0.20**	−0.16*	−0.12	−0.18*	−0.11	–					
11. Senses	−0.21**	−0.09	0.47**	−0.40**	−0.27**	−0.27**	−0.30**	−0.20**	−0.19**	0.04	–				
12. INV + 1	0.22**	0.00	−0.71**	0.43**	0.24**	0.17*	0.22**	0.25**	0.20**	−0.20**	−0.37**	–			
13. EE	−0.45**	−0.32**	0.09	−0.28**	0.04	0.01	−0.05	−0.05	−0.12	−0.26**	0.16*	−0.14	–		
14. Image	−0.12	−0.03	0.01	−0.24**	0.08	0.09	0.03	0.00	−0.05	0.09	0.13	0.01	0.49**	–	
15. BOI	0.09	0.28**	0.11	−0.26**	−0.01	−0.03	−0.01	−0.02	−0.10	0.23**	0.09	−0.11	0.30**	0.43**	–

**p* < 0.05, ***p* < 0.01.

CL, categorization latency; Freq, HAL log-frequency; AoA, age of acquisition; LOD, Levenshtein orthographic distance; Letters, number of letters; Phon, number of phonemes; Syll, number of syllables; Morph, number of morphemes; Conc, concreteness; Senses, number of senses; INV+ 1, inverse of number of word neighbors plus 1; EE, emotional experience; Image, imageability; BOI, body-object interaction.

experience dimension was in the opposite direction. These significant interactions mean that the regression lines for each dimension of experiential knowledge are not parallel (i.e., they have significantly different slopes) in the two SCTs. To better understand the precise nature of the effects of emotional experience, imageability, and BOI for each SCT, we conducted follow up hierarchical multiple regression analyses separately for each data set.

CONCRETE SCT

The hierarchical multiple regression results are shown in **Table 5**. (For both SCTs, the associated beta-weights and semi-partial correlations for the predictor variables are given only for the step at which they entered the multiple regression equation.) There are two important results that should be highlighted. First, at step 1 of the analyses, concreteness had significant negative semi-partial

Table 4 | Results of interaction tests in the omnibus analyses.

Variable	*B*	SEB	β	sr	ΔR^2	R^2
CATEGORIZATION LATENCY						
Step 1 (control variables)						0.35***
Step 2					0.14***	0.49***
Task × EE	−0.16	0.03	−0.81	−0.20***		
Task × imageability	0.25	0.05	0.66	0.18***		
Task × BOI	0.39	0.07	0.93	0.22***		
CATEGORIZATION ERROR						
Step 1 (control variables)						0.22***
Step 2					0.20***	0.42***
Task × EE	−4.05	0.64	−1.03	−0.25***		
Task × imageability	3.98	1.06	0.56	0.15***		
Task × BOI	10.46	1.35	1.30	0.31***		

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*
EE, emotional experience; BOI, body-object interaction.

Table 5 | Results of hierarchical multiple regression analyses for the concrete SCT.

Variable	*B*	*SEB*	β	*sr*	ΔR^2	R^2
CATEGORIZATION LATENCY						
Step 1						0.55***
Freq	−0.04	0.02	−0.20	−0.12*		
AoA	0.08	0.01	0.43	0.35***		
LOD	−0.11	0.04	−0.27	−0.14**		
Phonemes	0.00	0.02	0.00	0.00		
Syllables	0.03	0.05	0.04	0.03		
Morphemes	−0.02	0.04	−0.02	−0.02		
Concreteness	−0.35	0.05	−0.47	−0.37***		
Senses	0.03	0.01	0.13	0.11*		
NCOUNT−INV	−0.11	0.07	−0.12	−0.08		
Step 2					0.07***	0.62***
EE	0.03	0.03	0.06	0.06		
Imageability	−0.12	0.04	−0.19	−0.12**		
BOI	−0.13	0.03	−0.31	−0.24***		
CATEGORIZATION ERROR						
Step 1						0.39***
Freq	−0.66	0.38	−0.17	−0.10		
AoA	0.57	0.25	0.16	0.13*		
LOD	−2.43	0.83	−0.32	−0.17**		
Phonemes	−0.75	0.49	−0.18	−0.09		
Syllables	1.67	1.03	0.16	0.09		
Morphemes	−1.00	0.75	−0.09	−0.08		
Concreteness	−7.62	0.99	−0.57	−0.45***		
Senses	0.25	0.30	0.06	0.05		
NCOUNT-INV	−1.33	1.47	−0.08	−0.05		
Step 2					0.16***	0.55***
EE	1.85	0.51	0.21	0.18***		
Imageability	−2.75	0.86	−0.25	−0.16**		
BOI	−3.41	0.49	−0.46	−0.35***		

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*
Freq, HAL log-frequency; AoA, age of acquisition; LOD, Levenshtein orthographic distance; NCOUNT-INV, inverse of number of word neighbors plus 1; EE, emotional experience; BOI, body-object interaction.

correlations for both categorization latencies and categorization errors. That is, concreteness exerted a facilitatory effect, such that higher concreteness ratings (i.e., higher typicality ratings for the "concrete" category, see footnote 2) were associated with faster and more accurate categorizations (which is exactly what would be expected from "concrete" typicality ratings). The within-category variability accounted for by concreteness (or typicality) allowed for a more stringent test of the effects of the three dimensions of experiential knowledge at step 2 of the analyses. Second, and importantly, as can be seen in **Table 5**, for both criterion variables, there was a significant change in R^2 when the three dimensions of experiential knowledge were added to the analyses.

Recall that according to Vigliocco et al.'s (2009) framework of semantic representation, sensorimotor knowledge is diagnostic of concrete concepts. We therefore had predicted that imageability and BOI should exert facilitatory effects in the concrete SCT. These predictions were supported for both the categorization latency and categorization error data, such that higher imageability ratings and higher BOI ratings were associated with faster and more accurate categorizations. Our predictions regarding the effects of emotional experience were more speculative. One possibility we suggested was that because emotional experience is diagnostic of abstract concepts, it may inhibit categorization, because this dimension is not congruent with the decision criterion of "is the word concrete?". A second possibility we suggested was that this dimension may exert no effects (for the potential reasons outlined above). The data provided mixed support for the two predictions. There was an inhibitory effect of emotional experience on categorization errors, such that higher emotional experience ratings were associated with less accurate categorizations, although there was no effect of emotional experience on categorization latencies. These results will be examined in more detail in the Discussion section.

ABSTRACT SCT

The hierarchical multiple regression results are shown in **Table 6**. There are again two important results that should be highlighted. First, at step 1 of the analyses, concreteness had significant positive semi-partial correlations for both categorization latencies and categorization errors. That is, concreteness exerted a facilitatory effect, such that lower concreteness ratings (i.e., higher typicality ratings for the "abstract" category, see footnote 2) were associated with faster and more accurate categorizations (which is exactly what would be expected from "abstract" typicality ratings). Once more, the within-category variability accounted for by concreteness (or typicality) allowed for a more stringent test of the effects of the three dimensions of experiential knowledge at step 2 of the analyses. Second, and importantly, as can be seen in **Table 6**, for both criterion variables, there was a significant change in R^2 when the three dimensions of experiential knowledge were added to the analyses.

Recall that according to Vigliocco et al.'s (2009) framework of semantic representation, emotional knowledge is diagnostic of abstract concepts. We therefore had predicted that emotional experience should exert facilitatory effects. This prediction was supported for both the categorization latency and categorization error data, such that higher emotional experience ratings were associated with faster and more accurate categorizations. Our predictions regarding the effects of imageability and BOI were more speculative. One possibility we suggested was that because these two dimensions are diagnostic of concrete concepts, they may inhibit categorization, because they are not congruent with the decision criterion of "is the word abstract?". A second possibility we suggested was that these two dimensions may exert no effects (again, for the potential reasons outlined above). The data provided mixed support for the two predictions. On the one hand, BOI exerted inhibitory effects on categorization latencies and errors, such that higher BOI ratings were associated with slower and less accurate categorizations. On the other hand, imageability exerted no effect on either categorization latencies or errors. Again, these results will be examined in more detail in the Discussion section.

DISCUSSION

According to the PSS framework of grounded cognition, conceptual processing involves simulation, or the partial reenactment of the neural states involved during bodily interaction with the environment (Barsalou, 1999). More recently, Barsalou (2003, 2008, 2009) elaborated PSS to include the idea of situated conceptualization: representations underlying conceptual knowledge include much of the rich information associated with the environmental contexts in which those concepts were acquired. Thus, simulation of conceptual knowledge involves many forms of neural reenactment, such as sensory, motor, and emotional neural reenactment (Wilson-Mendenhall et al., 2011).

As mentioned, the PSS framework has previously been used to explain the effects of imageability and BOI in lexical processing. In conjunction with PSS, the semantic feedback activation framework has been used to provide a specific account for how effects of imageability and BOI arise within the visual word recognition system. The basic idea is that easily imageable words and high BOI words are semantically richer, and thus they generate greater amounts of semantic activation (i.e., richer sensory simulations and richer motor simulations) within semantic units. Facilitatory effects of imageability and BOI are observed in such tasks as lexical decision and word naming because the greater amount of semantic activation generated by easily imageable words and high BOI words leads to greater semantic feedback to orthographic units and to phonological units, which leads to faster settling of orthographic representations and phonological representations, respectively. Facilitatory effects of the above dimensions are observed in semantic categorization because the greater amount of semantic activation generated by easily imageable words and high BOI words leads to faster settling of semantic representations.

An important consideration when examining the effects of a particular semantic dimension, particularly for the present study in which the SCT was used, is that given the dynamic nature of semantic processing (e.g., Kiefer and Pulvermüller, 2012), the effect of any particular dimension is likely to be a function of both bottom-up processing (e.g., semantically richer words elicit greater levels of semantic activation) and the top-down influence of task demands. For example, the vast majority of studies examining the effects of BOI in the SCT have used imageability (e.g., Wellsby et al., 2011) or concreteness (Bennett et al., 2011) decision criteria. All these studies reported facilitatory effects of BOI. The explanation offered is that the increased semantic activation (or

Table 6 | Results of hierarchical multiple regression analyses for the abstract SCT.

Variable	*B*	*SEB*	β	*sr*	ΔR^2	R^2
CATEGORIZATION LATENCY						
Step 1						0.22***
Freq	−0.01	0.02	−0.06	−0.03		
AoA	0.01	0.01	0.05	0.04		
LOD	−0.05	0.04	−0.12	−0.08		
Phonemes	0.04	0.02	0.23	0.13		
Syllables	0.05	0.04	0.15	0.09		
Morphemes	−0.05	0.04	−0.11	−0.08		
Concreteness	0.12	0.03	0.31	0.30***		
Senses	−0.01	0.01	−0.08	−0.07		
NCOUNT−INV	0.11	0.06	0.16	0.11		
Step 2					0.15***	0.37***
EE	−0.11	0.02	−0.45	−0.35***		
Imageability	0.00	0.04	−0.01	−0.01		
BOI	0.16	0.06	0.19	0.16**		
CATEGORIZATION ERROR						
Step 1						0.17***
Freq	0.18	0.38	0.06	0.03		
AoA	0.00	0.24	0.00	0.00		
LOD	−0.55	0.94	−0.06	−0.04		
Phonemes	0.05	0.44	0.02	0.01		
Syllables	0.17	0.82	0.02	0.01		
Morphemes	0.31	0.81	0.04	0.03		
Concreteness	3.43	0.61	0.40	0.38***		
Senses	−0.37	0.29	−0.10	−0.09		
NCOUNT−INV	1.14	1.44	0.08	0.05		
Step 2					0.13***	0.30***
EE	−1.79	0.45	−0.33	−0.25***		
Imageability	−0.14	0.85	−0.01	−0.01		
BOI	6.11	1.35	0.34	0.29***		

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Freq, HAL log-frequency; AoA, age of acquisition; LOD, Levenshtein orthographic distance; NCOUNT-INV, inverse of number of word neighbors plus 1; EE, emotional experience; BOI, body-object interaction.

richer motor simulations) elicited by high BOI words provided evidence consistent with the demands of the task (e.g., respond only if the word is easily imageable or is concrete). However, Tousignant and Pexman (2012) explicitly manipulated the instructions given to their participants. More specifically, in three of their SCTs, participants knew that "entity" (concrete thing) was part of the decision category. In their first SCT, participants were instructed to press one button for words referring to entities and another button for words referring to non-entities. In their second and third SCTs, participants were told to press one button for words referring to entities and another for words referring to actions (the order of presentation in the instructions of which buttons to press for entity and action words were reversed in these two SCTs). In a fourth SCT, participants were instructed to press one button for words referring to actions and another for words referring to non-actions. Thus, in the fourth SCT, there was no explicit mention of entities in the instructions. Tousignant and Pexman (2012) reported facilitatory effects of BOI only for the three SCTs in which "entity" explicitly comprised part of the decision category. For these three SCTs, the increased semantic activation (or richer motor simulations) elicited by high BOI words provided evidence consistent with the demands of the task (e.g., respond in a specific way to words referring to entities). For the fourth SCT, the increased semantic activation (or richer motor simulations) elicited by high BOI words did not provide evidence consistent with the demands of the task, as it was essentially an "action versus no action" decision category, and thus no effect of BOI was observed. This consideration of the interaction between semantic activation and task demands will be important in our discussion below of the nature of the effects observed in the present study.

As noted, the purpose of the present study was to examine the effects of emotional experience, imageability, and BOI in semantic processing for concrete and abstract nouns. According to Vigliocco et al.'s (2009) framework of semantic representation, sensorimotor knowledge should underlie the meanings of concrete concepts, whereas emotional knowledge should underlie the meanings of abstract concepts. Based on this framework, we made two general sets of predictions, which we will address in turn.

First, we predicted that when any dimension of experiential knowledge is diagnostic of a type of concept examined in a particular SCT, the knowledge that dimension brings to bear should be congruent with the decision criterion of that SCT, and should thus lead to facilitation of task performance. What this means for the present study is that in the concrete SCT, because imageability and BOI are diagnostic of concrete concepts, higher ratings on these two dimensions should lead to faster and more accurate concreteness categorizations, whereas in the abstract SCT, because emotional experience is diagnostic of abstract concepts, higher ratings on this dimension should lead to faster and more accurate abstractness categorizations. All these predictions were supported, such that there were facilitatory effects of imageability and BOI in the concrete SCT (i.e., higher ratings on these two dimensions were associated with faster and more accurate concreteness categorizations), and there were facilitatory effects of emotional experience in the abstract SCT (i.e., higher ratings on this dimension were associated with faster and more accurate abstractness categorizations).

These results regarding the facilitatory effects of imageability and BOI in the concrete SCT and of emotional experience in the abstract SCT provide one important source of support for the idea that imageability and BOI are diagnostic of concrete concepts and that emotional experience is diagnostic of abstract concepts (Vigliocco et al., 2009). These results also strongly support the idea, derived from the semantic feedback activation framework of visual word recognition (Hino and Lupker, 1996; Pexman et al., 2002), that when a particular dimension of experiential knowledge is congruent with task demands, task performance is facilitated. In other words, these results are consistent with the literature outlined in the Introduction demonstrating that when a semantic richness variable provides evidence consistent with task demands, semantic categorization performance is facilitated. However, an important and novel aspect of the present study is that we observed facilitatory effects of a dimension of emotional experiential knowledge (i.e., the dimension of emotional experience) in the processing of nouns referring to abstract concepts.

Second, we predicted that two possible outcomes could occur when any dimension of experiential knowledge is not diagnostic of a type of concept examined in a particular SCT. One possible outcome was that inhibitory effects would be observed, and an alternative outcome was that no effects would be observed, under these experimental conditions. For the present study, this meant that either inhibitory or null effects of imageability and BOI were expected for the abstract SCT, whereas inhibitory or null effects of emotional experience were expected for the concrete SCT. The results did not provide unequivocal support for either prediction. In the concrete SCT, there was an inhibitory effect of emotional experience on categorization errors (higher ratings of emotional experience were associated with less accurate concreteness categorizations), but there was no effect on categorization latencies. In the abstract SCT, there were inhibitory effects of BOI on both categorization latencies and errors (higher ratings of BOI were associated with slower and less accurate abstractness categorizations), but there were no effects of imageability.

These inhibitory effects of BOI in the abstract SCT and of emotional experience in the concrete SCT provide a second source of support for the idea that motor knowledge is diagnostic of concrete concepts and that emotional knowledge is diagnostic of abstract concepts (Vigliocco et al., 2009). The reason for this is that when a knowledge type is not congruent with task demands (e.g., motor knowledge is not congruent with making abstractness categorizations), the increased levels of semantic richness (e.g., richer motor simulations) do not facilitate performance (e.g., making abstractness categorizations). This provides some support for the idea that simulation is an obligatory cognitive process, and is not simply used when it may facilitate performance (e.g., using motor simulations in the concrete SCT). However, we emphasize that these findings and conclusions are tentative because the results are novel, and future research will need to be undertaken to determine whether they are reliable (i.e., can be replicated), or are due to some theoretically uninteresting reason specific to the present study (e.g., the particular stimulus sets used).

In the present results, the concreteness (or, typicality) dimension was related to the processing of concrete nouns and abstract nouns. As noted in footnote 2, the higher end of the concreteness ratings (in the Friendly et al., 1982, and Paivio et al., 1968, norms) can be treated as measuring more typical instances of the category "concrete things," whereas the lower end of the concreteness ratings can be treated as measuring more typical instances of the category "abstract things." In the concrete SCT, there were significant negative semi-partial correlations between concreteness and categorization latency and categorization error. These findings indicate that higher concreteness ratings (i.e., higher typicality ratings of "concrete things") were associated with faster and more accurate categorizations. In the abstract SCT, there were significant positive semi-partial correlations between concreteness and categorization latency and categorization error. In this case, these findings indicate that lower concreteness ratings (i.e., higher typicality ratings of "abstract things") were associated with faster and more accurate categorizations. Thus, in both SCTs, categorization was facilitated for concepts rated to be more typical of the particular category (higher concreteness ratings for the concrete SCT, but lower concreteness ratings for the abstract SCT). Including concreteness in the analyses was important, because its inclusion allowed for more stringent tests of the effects of the three dimensions of experiential knowledge; any overlapping variability shared by these dimensions with concreteness was credited to concreteness at the first step of the analyses. Hence, any variability accounted for by emotional experience, imageability, and BOI in the present study is unique and not shared with typicality or any other of the measures included in the analyses.

An interesting question that the present study cannot address, due to the go/no-go nature of the two SCTs, is what would occur if a yes/no design were used (i.e., overt button responses are made to both items requiring "yes" responses and to items requiring "no" responses). More specifically, what would be the effects of emotional experience for abstract nouns presented on "no" trials in a concrete SCT, and what would be the effects of imageability and BOI for concrete nouns presented on "no" trials in an abstract SCT? Based on the semantic feedback activation framework of visual word recognition, the following predictions can be made. First, in the concrete SCT, because emotional experience is diagnostic of abstract concepts, lower ratings on this dimension would be associated with the noun being considered less abstract, or in other words, being considered more concrete, which would likely lead to inhibitory effects because these nouns would be more difficult to differentiate from concrete nouns. Second, in the abstract SCT, because imageability and BOI are diagnostic of concrete concepts, lower ratings on these two dimensions would be associated with the noun being considered less concrete, or in other words, being considered more abstract, which would likely lead to inhibitory effects because these nouns would be more difficult to differentiate from abstract nouns. Of course, these predictions must await testing in future research.

A final and important issue that was not directly addressed in the present study was how the dimension of emotional experience may be related to other dimensions of emotionality, such as valence and arousal, that have been used in the literature to assess the influence of emotional knowledge in lexical processing. Kousta et al. (2009, 2011) have demonstrated that valence and arousal significantly influence lexical processing in the lexical decision task. It is therefore important to examine how different measures of emotional experiential knowledge are related and of their effects in lexical processing.

To examine the specific issues of the relationships between the dimensions of emotional experience, valence, and arousal, and their effects on categorization latency and errors in the present study, we did the following. First, we obtained valence and arousal values from the Affective Norms for English Words (ANEW) database (Bradley and Lang, 1999), which were available for 87 of the concrete nouns and 69 of the abstract nouns. Second, we conducted separate *post hoc* simultaneous multiple regression analyses for each data set. (We conducted simultaneous regression analyses rather than hierarchical regression analyses because of the reduction of statistical power due to the smaller number of stimuli in each analysis.) All the lexical and semantic variables that were entered in the follow up analyses above were entered in the *post hoc* analyses, along with valence and arousal. We emphasize that these analyses are exploratory in nature, and any conclusions that may be derived from them are tentative and must await further experimentation.

For the concrete nouns, the zero-order correlations between valence and arousal, valence and emotional experience, and arousal and emotional experience were $r(87) = 0.34$ ($p < 0.01$), 0.41 ($p < 0.001$), and 0.57 ($p < 0.001$), respectively. For the abstract nouns, the zero-order correlations between valence and arousal, valence and emotional experience, and arousal and emotional experience were $r(69) = 0.11$ (*ns*), -0.08 (*ns*), and 0.60 ($p < 0.001$), respectively. These correlations suggest that the dimension of emotional experience is positively related to the dimension of arousal for both concrete nouns and abstract nouns, whereas it is only positively related to the dimension of valence for the concrete nouns. The positive relationship between emotional experience and valence is perhaps not surprising, considering that valence was an emotional characteristic that is salient in the instructions used to obtain the emotional experience ratings.

For the concrete SCT regression analyses, although none of the three dimensions of emotional experiential knowledge were significantly related to categorization latencies, valence was significantly related to categorization errors, such that higher ratings of valence were associated with less accurate categorization ($sr = 0.15$). For the abstract SCT regression analyses, only emotional experience was significantly related to categorization latency, such that higher ratings of emotional experience were associated with faster latencies ($sr = -0.39$), and none of the dimensions of emotional experiential knowledge were significantly related to categorization errors. The most important finding from the *post hoc* regression analyses was that although emotional experience and arousal were significantly positively correlated for the abstract noun stimulus set, emotional experience continued to exert a facilitatory effect on categorization latencies in the abstract SCT, even with arousal in the analysis. This finding provides further support for the idea that the dimension of emotional experience is a robust measure of emotional experiential knowledge. Of course, further research is needed to determine how this dimension is related to other dimensions of emotional experiential knowledge, and of their influence in visual word recognition tasks other than semantic categorization.

REFERENCES

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. J. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. Gen.* 133, 283–316.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Lang. Cogn. Process.* 18, 513–562.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Phil. Trans. R. Soc. B* 364, 1281–1289.

Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2008). "Language and simulation in conceptual processing," in *Symbols, Embodiment, and Meaning*, eds M. De Vega, A. M. Glenberg, and A. C. Graesser (Oxford, UK: Oxford University Press), 245–284.

Bennett, S. D. R., Burnett, A. N., Siakaluk, P. D., and Pexman, P. M. (2011). Imageability and body-object interaction ratings for 599 multisyllabic nouns. *Behav. Res. Methods* 43, 1100–1109.

Bradley, M. M., and Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings (Technical Report C-1)*. Gainesville, FL: Center for Research in Psychophysiology, University of Florida.

Casey, P. J. (1992). A reexamination of the roles of typicality and category dominance in verifying category membership. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 823–834.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science* 298, 1191–1194.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge: MIT Press.

Friendly, M., Franklin, P. E., Hoffman, D., and Rubin, D. C. (1982). The Toronto Word Pool: norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behav. Res. Methods Instrum. Comput.* 14, 375–399.

Griffiths, T. L., Steyvers, M., and Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychol. Rev.* 114, 211–244.

Hampton, J. A. (1997). Associative and similarity-based processes in categorization decisions. *Mem. Cognit.* 25, 625–640.

Hansen, D., Siakaluk, P. D., and Pexman, P. M. (2012). The influence of print exposure on the body-object interaction effect in visual word recognition. *Front. Hum. Neurosci.* 6:113. doi:10.3389/fnhum.2012.00113

Hargreaves, I. S., Leonard, G., Pexman, P. M., Pittman, D., Siakaluk, P. D., and Goodyear, B. G. (2012). The neural correlates of the body-object interaction effect in semantic processing. *Front. Hum. Neurosci.* 6:22. doi:10.3389/fnhum.2012.00022

Hino, Y., and Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: an alternative to lexical access accounts. *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1331–1356.

Keith, T. Z. (2006). *Multiple Regression and Beyond.* Boston: Allyn and Bacon.

Kiefer, M., and Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex* 48, 805–825.

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., and Del Campo, E. (2011). The representation of abstract words: why emotion matters. *J. Exp. Psychol. Gen.* 140, 14–34.

Kousta, S.-T., Vinson, D. P., and Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition* 112, 473–481.

Kuperman, V., Stadthagen-Gonzalez, H., and Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behav. Res. Methods.* doi: 10.3758/s13428-012-0210-4

Landauer, T., and Dumais, S. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychol. Rev.* 104, 211–240.

Larochelle, S., and Pineau, H. (1994). Determinants of response times in the semantic verification task. *J. Mem. Lang.* 33, 796–823.

Lund, K., and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods* 28, 203–208.

Niedenthal, P. M., Barsalou, L. W., Ric, F., and Krauth-Gruber, S. (2005a). "Embodiment in the acquisition and use of emotion knowledge," in *Emotion and Consciousness* eds L. Feldman Barrett, P. M. Niedenthal, and P. Winkielman Emotion, and consciousness (New York: Guilford), 21–50.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., and Ric, F. (2005b). Embodiment in attitudes, social perception, and emotion. *Pers. Soc. Psychol. Rev.* 9, 184–211.

Niedenthal, P. M., Winkielman, P., Mondillon, L., and Vermeulen, N. (2009). Embodiment of emotion concepts. *J. Pers. Soc. Psychol.* 96, 1120–1136.

Paivio, A., Yuille, J. C., and Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *J. Exp. Psychol.* 76, 1–25.

Parisi, D. (2011). The other half of the embodied mind. *Front. Psychol.* 2:69. doi:10.3380/fpsyg.2011.00069

Pecher, D., and Zwaan, R. A. (eds). (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking.* Cambridge, UK: Cambridge University Press.

Pexman, P. M., Holyk, G. G., and Monfils, M.-H. (2003). Number of features effects and semantic processing. *Mem. Cognit.* 31, 842–855.

Pexman, P. M., Lupker, S. J., and Hino, Y. (2002). The impact of feedback semantics in visual word recognition: number of features effects in lexical decision and naming tasks. *Psychon. Bull. Rev.* 9, 542–549.

Phillips, C. I., Sears, C. R., and Pexman, P. M. (2012). An embodied semantic processing effect on eye gaze during sentence reading. *Lang. Cogn.* 4, 99–114.

Pylyshyn, Z. (1984). *Computation and Cognition: Towards a Foundation for Cognitive Science.* Cambridge: MIT Press.

Shaoul, C., and Westbury, C. (2010a). Exploring lexical co-occurrence space using HiDEx. *Behav. Res. Methods* 42, 393–413.

Shaoul, C., and Westbury, C. (2010b). *Neighborhood Density Measures for 57,153 English words.* Edmonton, AB: University of Alberta (downloaded from http://www.psych.ualberta.ca/~westburylab/downloads/westburylab.arcs.ncounts.html)

Siakaluk, P. D., Pexman, P. M., Aguilera, L., Owen, W. J., and Sears, C. R. (2008a). Evidence for the activation of sensorimotor information during visual word recognition: the body-object interaction effect. *Cognition* 106, 433–443.

Siakaluk, P. D., Pexman, P. M., Sears, C. R., Wilson, K., Locheed, K., and Owen, W. J. (2008b). The benefits of sensorimotor knowledge: body-object interaction facilitates semantic processing. *Cogn. Sci.* 32, 591–605.

Smith, E. E., Shoben, E. J., and Rips, L. J. (1974). Structure and process in semantic memory: a feature model for semantic decisions. *Psychol. Rev.* 81, 214–241.

Tillotson, S. M., Siakaluk, P. D., and Pexman, P. M. (2008). Body-object interaction ratings for 1,618 monosyllabic nouns. *Behav. Res. Methods* 40, 1075–1078.

Tousignant, C., and Pexman, P. M. (2012). Flexible recruitment of semantic richness: context modulates body-object interaction effects in lexical-semantic processing. *Front. Hum. Neurosci.* 6:53. doi:10.3389/fnhum.2012.00053

Vigliocco, G., Meteyard, L., Andrews, M., and Kousta, S. (2009). Toward a theory of semantic representation. *Lang. Cogn.* 1, 219–248.

Wellsby, M., Siakaluk, P. D., Owen, W. O., and Pexman, P. M. (2011). Embodied semantic processing: the body-object interaction effect in a non-manual task. *Lang. Cogn.* 3, 1–14.

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., and Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia* 49, 1105–1127.

Yap, M. J., Pexman, P. M., Wellsby, M., Hargreaves, I. S., and Huff, M. (2012). An abundance of riches: cross-task comparisons of semantic richness effects in visual word recognition. *Front. Hum. Neurosci.* doi:10.3389/fnhum.2012.00053

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 July 2012; accepted: 19 September 2012; published online: 08 October 2012.

Citation: Newcombe PI, Campbell C, Siakaluk PD and Pexman PM (2012) Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. Front. Hum. Neurosci. ***6****:275. doi: 10.3389/fnhum.2012.00275*

APPENDIX

CONCRETE NOUNS USED IN THE EXPERIMENTS

*accordion
acid
agent
alligator
aluminum
ambassador
apple
arrow
artist
asparagus
auditorium
author
baby
bacteria
barrel
basement
basin
bedroom
berry
binoculars
blacksmith
blanket
body
building
bullet
*business
butcher
butter
cabin
canal
candy
cannon
canoe
captain
cardinal
carpet
carriage
castle
caterpillar
cattle
cauliflower
chairman
chapter
cherry
chestnut
chickenpox
chimney
chinchilla
chopstick
cigarette
circle
city
clarinet
*climate
closet
clothing
coffee
collar
college
colonel
color
column
compass
comrade
concert
contract
corner
costume
cotton
country
couple
cousin
creature
crocodile
crystal
cupboards
dandruff
daughter
daylight
dealer
*destroyer
diamond
dinner
*disease
dishwasher
dollar
doorway
dragon
eagle
earthworm
elbow
empire
engine
envelope
estate
*evening
fabric
farmer
finger
flashbulb
football
forehead
forest
fountain
freckles
garden
grasshopper
handlebars
headboard
helmet
highway
hotel
hunter
hurricane
husband
*illness
island
jacket
jellyfish
jewel
jitterbug
journal
kitchen
lady
*laughter
leader
leather
letter
lieutenant
lion
luncheon
machine
madam
marshmallow
mayor
*meeting
member
merchant
metal
mirror
mistress
*moisture
monarch
monastery
money
monster
mother
motor
mountain
mouthpiece
navy
number
orchard
oven
painter
painting
partner
pasture
penicillin
perfume
person
picture
pitcher
planet
platform
player
poem
police
pony
prairie
projector
propeller
province
quarter
railway
rattlesnake
rectangle
refrigerator
sandwich
scholar
screwdriver
secretary
sentence
servant
shepherd
sheriff
sparrow
speaker
squirrel
station
steamer
stomach
student
summer
thermometer
trapezoid
treasurer
tuberculosis
window
winter

ABSTRACT NOUNS USED IN THE EXPERIMENTS

aberration	contrast	fallacy	*jeopardy	prestige
ability	control	fantasy	judgment	quality
absence	courage	fate	*justice	rating
accord	crisis	favor	knowledge	reaction
*account	criterion	feature	legend	reason
advance	custom	feeling	limit	reform
adversity	danger	feint	maker	regard
advice	deceit	feudalism	malice	relief
afterlife	decline	figment	manner	request
agreement	decrease	folly	marvel	reserve
allegory	deduction	forethought	mastery	revenge
amount	degree	fortune	meaning	review
appeal	delay	freedom	meantime	satire
approach	democracy	future	memory	sensation
aptitude	desire	gist	menace	*sister
array	devotion	gratitude	mercy	situation
aspect	*dijon	greed	merit	sobriety
atrocity	discipline	habit	method	soul
attempt	disclosure	heredity	mind	spirit
attitude	discretion	hindrance	miracle	status
attribute	disposition	*honor	*moment	support
banality	distraction	hope	mood	suppression
basis	distress	*hour	necessity	suspect
belief	duty	hypothesis	neglect	temerity
betrayal	eccentricity	idea	non-sense	tendency
blandness	economy	ignorance	nothing	theory
boredom	effect	illusion	notion	tribute
capacity	effort	*image	obedience	*trifle
chance	emancipation	immunity	obsession	trouble
clemency	envy	impulse	offense	truth
comment	equity	inanity	offer	unification
comparison	error	incident	opinion	upkeep
competence	essence	incline	opportunity	value
*compound	exclusion	inducement	outcome	vanity
concept	export	ingratitude	pacifism	venture
concern	expression	instance	pardon	violation
confidence	extent	instant	patience	virtue
conflict	*facility	intellect	perception	weakness
consent	factor	interest	perjury	welfare
context	failure	irony	pledge	wonder

WRITTEN INSTRUCTIONS USED FOR THE EMOTIONAL EXPERIENCE RATING TASK

Words differ in the extent to which they elicit or evoke an emotional experience. Some words elicit or evoke strong emotional experiences (e.g., JUSTICE), whereas other words elicit or evoke weaker emotional experiences (e.g., MOMENT). The purpose of this experiment is to rate words as to the ease with which they elicit or evoke emotional experience. For example, the word "justice" refers to a concept that is associated with high levels of emotional experience (e.g., think of the emotional conditions that arise when a jury verdict is delivered, such as joy, dismay, anger, frustration), whereas the word "moment" refers to a concept that is associated with low levels of emotional experience (i.e., it is difficult to think of any kind of emotional experience to which this word is related). Any word (e.g., "justice") that in your estimation elicits or evokes high levels of emotional experience should be given a high emotional experience rating (at the upper end of the numerical scale). Any word (e.g., "moment") that in your estimation elicits or evokes low levels of emotional experience should be given a low emotional experience rating (at the lower end of the scale). Because words tend to make you think of other words as associates, it is important that your ratings *not* be based on this and that you judge only the ease with which a word elicits or evokes emotional experience. Remember, all the words are nouns and you should base your ratings on this fact.

Your emotional experience ratings will be made on a 1–7 scale. A value of 1 will indicate a low emotional experience rating, and a value of 7 will indicate a high emotional experience rating. Values of 2–6 will indicate intermediate ratings. Please feel free to use the whole range of values provided when making your ratings. Circle the rating that is most appropriate for each word. When making your ratings, try to be as accurate as possible, but do not spend too much time on any one word.

1	2	3	4	5	6	7
Low			Medium			High