9

INDIVIDUAL DIFFERENCES IN SEMANTIC PRIMING PERFORMANCE

Insights from the Semantic Priming Project

Melvin J. Yap
Keith A. Hutchison
Luuan Chin Tan

Abstract

The *semantic/associative priming effect* refers to the finding of faster recognition times for words preceded by related targets (e.g. *cat*—*DOG*), compared to words preceded by unrelated targets (e.g. *hat*—*DOG*). Over the past three decades, a voluminous literature has explored the influence of semantic primes on word recognition, and this work has been critical in shaping our understanding of lexical processing, semantic representations, and automatic versus attentional influences. That said, the bulk of the empirical work in the semantic priming literature has focused on group-level performance that averages across participants, despite compelling evidence that individual differences in reading skill and attentional control can moderate semantic priming performance in systematic and interesting ways. The present study takes advantage of the power of the semantic priming project (SPP; Hutchison et al., 2013) to answer two broad, related questions. First, how stable are semantic priming effects, as reflected by within-session reliability (assessed by split-half correlations) and between-session reliability (assessed by test–retest correlations)? Second, assuming that priming effects *are* reliable, how do they interact with theoretically important constructs such as reading ability and attentional control? Our analyses replicate and extend earlier work by Stolz, Besner, & Carr (2005) by demonstrating that the reliability of semantic priming effects strongly depends on prime-target association strength, and reveal that individuals with more attentional control and reading ability are associated with stronger priming.

Words preceded by a semantically related word (e.g. *cat*—*DOG*) are recognized faster than when preceded by a semantically unrelated word (e.g. *hat*—*DOG*) (Meyer & Schvaneveldt, 1971). This deceptively simple phenomenon, which has been termed the *semantic priming effect*, is one of the most important observations in cognitive science, and has profoundly shaped our understanding of word recognition processes, the nature of the semantic system, and the distinction between automatic and controlled processes (for excellent reviews, see McNamara, 2005; Neely, 1991). Priming phenomena are too complex to be explained by a single process, and various theoretical mechanisms, varying from *automatic* (i.e. conscious awareness not required) to *strategic* (i.e. controlled and adaptively modulated by task context), have been invoked to explain priming (Neely, 1991). Priming mechanisms can also operate *prospectively* (i.e. before a target is presented) or *retrospectively* (i.e. after a target is presented).

The most well-known prospective mechanism is *automatic spreading activation* (Posner & Snyder, 1975). That is, a prime (e.g. *cat*) automatically pre-activates related nodes (e.g. DOG) via associative/semantic pathways, facilitating identification of these words when they are subsequently presented. Priming effects also seem to reflect controlled processes such as *expectancy* and *semantic matching*. Expectancy operates prospectively and refers to the intentional generation of candidates for the to-be-presented target (Becker, 1980), whereas semantic matching reflects a retrospective process that searches for a relation from the target to the prime (Neely, Keefe, & Ross, 1989). The presence of a prime-target relationship is diagnostic of the target's lexical status, because non-words cannot be related to their primes.

In addition to spreading activation, expectancy, and semantic matching, researchers have also proposed hybrid mechanisms that combine automatic and strategic aspects. For example, Bodner & Masson's (1997) *memory-recruitment* account of priming, which is itself based on Whittlesea & Jacoby's (1990) retrieval account of priming, is predicated on the idea that the processing of the prime establishes an episodic memory trace that can then be used to facilitate processing of the upcoming target. Importantly, the extent to which the system relies on the episodic trace of the prime is dependent on the prime's task relevance (Anderson & Milson, 1989). Interestingly, memory recruitment processes are postulated to operate even when primes are presented too briefly to be consciously processed (but see Kinoshita, Forster, & Mozer, 2008; Kinoshita, Mozer, & Forster, 2011, for an alternative account of Bodner & Masson's, 1997, findings). Of course, we should clarify that the foregoing mechanisms are not mutually exclusive and very likely operate collectively to produce priming.

For the most part, the semantic priming literature has focused on group-level data, which are collapsed across participants. Likewise, most models of semantic priming have not taken into account the systematic individual differences that exist among skilled readers (but see Plaut & Booth, 2000). However, this pervasive emphasis on the characterization of a "prototypical" reader is difficult to reconcile with mounting evidence that readers vary on dimensions which moderate word recognition performance (see Andrews, 2012, for a review). Are individual differences in the magnitude of a participant's semantic priming effect related to individual differences such as reading comprehension ability, vocabulary size, and attentional control? Closely connected to this question is whether semantic priming effects are reliable. More specifically, to what extent is the magnitude of an individual's semantic priming effect stable *within* and *across* experimental sessions? Should it turn out that semantic priming is inherently unreliable (e.g. Stolz, Besner, & Carr, 2005), this will severely limit the degree to which the magnitude of an individual's semantic priming effect can be expected to correlate with other measures of interest (Lowe & Rabbitt, 1998). Related to this, an unreliable dependent measure makes it harder for researchers to detect between-group differences on this measure (Waechter, Stolz, & Besner, 2010).

The present study capitalizes on the power of the SPP (Hutchison et al., 2013) to address the above-mentioned questions. The SPP is a freely accessible online repository (http://spp.montana.edu) containing lexical and associative/semantic characteristics for 1,661 words, along with *lexical decision* (i.e. classify letter strings as words or non-words, e.g. *flirp*) and *speeded pronunciation* (i.e. read words aloud) behavioral data of 768 participants from four testing universities (512 in lexical decision and 256 in speeded pronunciation). Data were collected over two sessions, separated by no more than one week. Importantly, Hutchison et al. (2013) also assessed participants on their attentional control (Hutchison, 2007), vocabulary knowledge, and reading comprehension. It is noteworthy that the SPP contains data for over 800,000 lexical decision trials and over 400,000 pronunciation trials, collected from a large and diverse sample of participants, making this a uniquely valuable resource for studying individual differences in primed lexical processing.

The SPP exemplifies the *megastudy* approach to studying lexical processing, in which researchers address a variety of questions using databases that contain behavioral data and lexical characteristics for very large sets of words (see Balota, Yap, Hutchison, & Cortese, 2012, for a review). Megastudies allow the language to define the stimuli, rather than compelling experimenters to select stimuli based on a limited set of criteria. They now serve as an important complement to traditional factorial designs, which are associated with selection artifacts, list context effects, and limited generalizability (Balota et al., 2012; Hutchison et al., 2013). For example, in the domain of semantic priming, it is often methodologically challenging to examine differences in priming as a function of some other categorical variable (e.g. target word frequency), due to the difficulty of matching experimental conditions on the many dimensions known to influence word recognition. Using megastudy data, regression analyses can be conducted to examine the effects of item characteristics on priming, with other correlated variables statistically controlled for (e.g. Hutchison, Balota, Cortese, & Watson, 2008).

We will be using the SPP to address two broad and related questions. First, how stable are semantic priming effects, as reflected by within-session reliability (assessed by split-half correlations) and between-session reliability (assessed by test–retest correlations)? Second, assuming that priming effects *are* stable, how are they moderated by theoretically important constructs such as reading ability and attentional control? While these questions are not precisely novel, they have received relatively little attention in the literature. To our knowledge, the present study is the first attempt to answer these intertwined questions in a unified and comprehensive manner, using an unusually large and well-characterized set of items and participants. We will first provide a selective review of studies that have explored individual differences in semantic priming, before turning to an important study by Stolz et al. (2005), who were the first to explore the reliability of semantic priming effects in word recognition.

Individual Differences in Semantic Priming

The great majority of the word recognition literature has been dominated by what Andrews (2012) calls the *uniformity assumption*. Specifically, there seems to be an implicit assumption that the architecture of the lexical processing system is relatively invariant across skilled readers, and this assumption is reflected in the field's reliance on measures of group-level performance which aggregate across relatively small samples (typically 25–35) of participants. However, even within a sample of college students, who are already selected for their reading and writing ability, there remains substantial variability in psychometric measures of reading and spelling, and in experimental measures of word recognition performance (Andrews & Hersch, 2010).

There is considerable evidence against the uniformity hypothesis, predominantly from word recognition studies where participants identify words presented in isolation (i.e. words are not preceded by a prime). For example, two important aspects of reading ability are *vocabulary knowledge* (i.e. knowledge of word forms and meaning) and *print exposure* (i.e. amount of text read). In both isolated lexical decision and speeded pronunciation, there is evidence that participants high on vocabulary knowledge and print exposure recognize words faster and more accurately, and are generally less influenced by lexical characteristics such as *word frequency*, *word length* (i.e. number of letters), and *number of orthographic neighbors* (i.e. words obtained by substituting a single letter in the target word, e.g. *sand*'s neighbors include *band* and *sang*) (Butler & Hains, 1979; Chateau & Jared, 2000; Lewellen, Goldinger, Pisoni, & Greene, 1993). However, because processing time is positively correlated with the size of experimental effects (Faust, Balota, Spieler, & Ferraro, 1999), it is possible that skilled lexical processors are showing smaller effects of stimulus characteristics simply because they are responding faster. Yap, Balota, Sibley, and Ratcliff (2012), in their large-scale analysis of data from the English Lexicon Project (Balota et al., 2007; http://elexicon.wustl.edu), controlled for variability in participants' processing speed by using z-score standardized response times (RTs), and still found that participants with more vocabulary knowledge produced smaller effects of lexical variables. These findings are consistent with the general perspective that as readers acquire more experience with written language, they become more reliant on relatively automatic lexical processing mechanisms, and are consequently less influenced by word characteristics (Yap et al., 2012).

To a lesser degree, researchers have also used priming paradigms to explore the influence of individual differences on word recognition. In the typical priming experiment, two letter strings (which are related in some manner) are presented to the participants, with the first letter string serving as the prime and the second as the target. Strings might be morphologically (*touching*—*TOUCH*), orthographically (*couch*—*TOUCH*), phonologically (*much*—*TOUCH*), or semantically (*feel*—*TOUCH*) related; researchers also have the option of masking the prime (i.e. presenting it too briefly to be consciously processed) to shed light on early lexical processes and to minimize strategic effects (Forster, 1998). A small number of studies have examined how priming effects are moderated by individual differences. For example, Yap, Tse, and Balota (2009) examined the joint effects of semantic relatedness (related versus unrelated prime) and target frequency (high versus low) on lexical decision performance, and found that these joint effects were moderated by the vocabulary knowledge of the participants. Specifically, participants with more vocabulary knowledge produced additive effects of priming and frequency (i.e. equal priming for low and high-frequency words), whereas participants with less vocabulary knowledge produced an overadditive interaction (i.e. stronger priming for low, compared to high-frequency, words).

These results are consistent with the idea that for high vocabulary knowledge participants, both high and low-frequency words are associated with strong, fully specified lexical representations which can be fluently accessed. Analyses of RT distributions revealed that for such participants, semantic priming largely reflects a relatively modular "head-start" mechanism, whereby primes prospectively pre-activate related high and low-frequency targets to a similar extent through spreading activation or expectancy-based processes, thereby speeding up lexical access by some constant amount of time (Yap et al., 2009). In contrast, participants with less vocabulary knowledge possess weaker lexical representations, particularly for low-frequency words. For these participants, head-start mechanisms still contribute to priming, but there is increased retrospective reliance on prime information for low-frequency words, which yields more priming for such words.

In addition to vocabulary knowledge, there is intriguing evidence that semantic priming mechanisms are modulated by individual differences in *attentional control* (AC). Broadly speaking, AC refers to the capacity to coordinate attention and memory so as to optimize task performance by enhancing task-relevant information (Hutchison, 2007), especially when the environment is activating conflicting information and prepotent responses (Shipstead, Lindsey, Marshall, & Engle, 2014). A number of studies have demonstrated that when the stimulus onset asynchrony (SOA) between the prime and target is sufficiently long for participants to form expectancies, an increase in *relatedness proportion* (i.e. the proportion of word prime-word target pairs that are related) leads to an increase in priming (see Hutchison, 2007, for a review); this is known as the *relatedness proportion effect*. By definition, increasing relatedness proportion increases the validity of the prime, leading participants to rely more heavily on expectancy-based processes, which in turn increases priming. Hutchison (2007) demonstrated that participants who were associated with greater AC (as reflected by performance on a battery of attentionally demanding tasks) produced a larger relatedness proportion effect, suggesting that high-AC participants were sensitive to contextual variation in relatedness proportion, and were adaptively increasing their reliance on expectancy generation processes as prime validity increased. Related to this, there is also recent evidence, based on the use of *symmetrically associated* (e.g. *sister—BROTHER*), *forward associated* (i.e. prime to target; e.g. *atom—BOMB*), and *backward associated* (i.e. target to prime; e.g. *fire—BLAZE*) prime-target pairs that high-AC individuals are more likely to prospectively generate expected targets and hold them in working memory, whereas low-AC individuals are more likely to engage retrospective semantic matching processes (Hutchison, Heap, Neely, & Thomas, 2014).

In summary, the individual differences literature is consistent with the idea that as readers accrue more experience with words, they rely more on automatic lexical processing mechanisms in both isolated (Yap et al., 2012) and primed (Yap et al., 2009) word recognition. This is consistent with the *lexical quality hypothesis* (Perfetti & Hart, 2001), which says that highly skilled readers are associated with high-quality lexical representations which are both fully specified and redundant. For these skilled readers, the process of identifying a word involves the precise activation of the corresponding underlying lexical presentation, with minimal activation of orthographically similar words (Andrews & Hersch, 2010). Furthermore, such readers are less dependent on the strategic use of context (e.g. prime information) to facilitate lexical retrieval (Yap et al., 2009). In addition to high-quality representations, lexical processing is also modulated by the extent to which readers can exert attentional control; high-AC readers are able to flexibly adjust their reliance on priming mechanisms so as to maximize performance on a given task (see Balota & Yap, 2006).

Is Semantic Priming Reliable?

Reliability, which is typically assessed in the domain of psychological testing, is a fundamental psychometric property that reflects the stability or consistency of a measure. The consistency of a measure can be evaluated across time (test–retest reliability) and across items within a test (split-half reliability). Establishing reliability is a critical prerequisite in the development of paper-and-pencil measures of intelligence, aptitude, personality, interests, and attitudes, but relatively little attention has been paid to the reliability of RT measures (Goodwin, 2009). Without first establishing reliability, one cannot tell if variability on a measure reflects meaningful individual differences or measurement noise. In order to address this, Yap et al. (2012), using trial-level data from the English Lexicon Project (Balota et al., 2007), examined the stability of word recognition performance in isolated lexical decision and speeded pronunciation. They found reassuringly high within and between-session correlations in the means and standard deviations of RTs, as well as RT distributional characteristics across distinct sets of items, suggesting that participants are associated with a relatively stable processing profile that extends beyond simple mean RT (see also Yap, Sibley, Balota, Ratcliff, & Rueckl, in press). There was also reliability in participants' sensitivity to underlying lexical dimensions. For example, participants who produced large word frequency effects in session 1 also tended to produce large frequency effects in session 2.

However, while the findings above lend support to the idea that participants respond to the influence of item *characteristics* in a reliable manner, it is less clear if the influence of item *context* is equally reliable. In other words, are semantic priming effects stable in the way word frequency or word length effects are? The answer to this question is surprisingly inconclusive. In semantic priming studies, there is tremendous variability in the magnitude of priming produced across different participants (Stolz et al., 2005). In order to determine if these individual differences reflect systematic or random processes, Stolz and colleagues (2005) examined the split-half and test–retest reliability of semantic priming across different experimental conditions where relatedness proportion and SOA were factorially manipulated. Surprisingly, they found that reliability was zero under certain conditions (e.g. short SOA, low relatedness proportion), which maximized the impact of automatic priming mechanisms and became statistically significant only under conditions (e.g. long SOA, high relatedness proportion), which made strategically mediated priming more likely. According to Stolz et al. (2005), these results point to a semantic system whose activity is "inherently noisy and uncoordinated" (p. 328) when priming primarily reflects automatic spreading activation, with performance becoming more coherent when task demands increase the controlled influence of mechanisms such as expectancy.

Such an observation places major constraints on researchers who intend to use semantic priming as a tool to study individual differences in domains such as personality (e.g. Matthews & Harley, 1993) or psychopathology (e.g. Morgan, Bedford, & Rossell, 2006), or who are interested in exploring how semantic priming might be moderated by variables such as attentional control (Hutchison, 2007), vocabulary knowledge (Yap et al., 2009), and item characteristics (Hutchison et al., 2008).

The Present Study

As mentioned earlier, we intend to leverage on the power of the SPP (Hutchison et al., 2013) to address the related questions of reliability and individual differences in semantic priming. Stolz et al. (2005) were the first to report the unexpectedly low reliabilities associated with the semantic priming effect. In Stolz et al. (2005), the number of participants in each experiment ranged between 48 and 96, priming effects were based on 25 related and 25 unrelated trials, and the two test sessions were operationalized by presenting two blocks of trials within the same experiment. Given the theoretical and applied importance of Stolz et al.'s (2005) findings, it is worthwhile exploring if they generalize to the SPP sample, which contains a much larger number of participants and items. Specifically, in the SPP, 512 participants contributed to the lexical decision data, priming effects are based on at least 100 related and 100 unrelated trials, and the two test sessions were held on different days, separated by no more than one week. The SPP was also designed to study priming under different levels of SOA (200 ms versus 1200 ms) and prime type (first associate versus other associate), allowing us to assess reliability across varied priming contexts.

Of course, for our purposes, reliability is largely a means to an end. The other major aspect of the present work concerns the extent to which a participant's semantic priming effect is predicted by theoretically important measures of individual differences, including vocabulary knowledge, reading comprehension, and attentional control. Although there has been some work relating priming to vocabulary knowledge (Yap et al., 2009) and to attentional control (Hutchison, 2007), there has not been, to our knowledge, a systematic exploration of the relationship between semantic priming effects and a comprehensive array of individual differences measures. The present study will be the first to address that gap, by considering how priming effects that are assessed to be reliable are moderated by a host of theoretically important variables. Collectively, the results of these analyses will help shed more light on the relationships between the quality of underlying lexical representations, attentional control, and priming phenomena. Our findings could also be potentially informative for more foundational questions pertaining to how changes in reading ability are associated with developments in the semantic system (e.g. Nation & Snowling, 1999).

Method

Dataset

All analyses reported in this chapter are based on archival trial-level lexical decision data from the semantic priming project (see Hutchison et al., 2013, for a full description of the dataset). The 512 participants were native English speakers recruited from four institutions (both private and public) located across the midwest, northeast, and northwest regions of the United States. Data were collected over two sessions on different days, separated by no more than one week. Across both sessions, participants received a total of 1,661 lexical decision trials (half words and half non-words), with word prime–word target pairs selected from the Nelson, McEvoy, & Schreiber (2004) free association norms; the relatedness proportion was fixed at 0.50. Non-words were created from word targets by changing one or two letters of each word to form pronounceable non-words that did not sound like real words. For each participant, each session comprised two blocks (a 200-ms and 1200-ms SOA block), and within each block, half the related prime-target pairs featured *first associates* (i.e. the target is the most common response to a cue/prime word, e.g. *choose*—*PICK*) while the other half featured *other associates* (i.e. any response other than the most common response to a cue, e.g. *preference*—*PICK*).

Additional demographic information collected included performance on the vocabulary and passage comprehension subtests of the Woodcock–Johnson III diagnostic reading battery (Woodcock, McGrew, & Mather, 2001) and on Hutchison's (2007) attentional control battery. The vocabulary measures include a synonym, antonym, and an analogy test; for reading comprehension, participants have to read a short passage and identify a missing keyword that makes sense in the context of that passage. The attentional control battery consists of an automated operational span (Unsworth, Heitz, Schrock, & Engle, 2005), Stroop, and antisaccade task (Payne, 2005). In the operational span task, participants have to learn and correctly recall letter sequences while solving arithmetic problems. In the Stroop task, participants are presented with incongruent (e.g. *red* printed in green), congruent (e.g. *green* printed in green), and neutral (e.g. *deep* printed in green) words and are required to name the ink color of the word as quickly and accurately as possible; the dependent variable is the difference in the mean RT or error rate between the congruent and incongruent conditions. In the antisaccade task, participants are instructed to look away from a flashed start (*) in order to identify a target (O or Q) that is briefly presented on the other side of the screen; the dependent variable is the target identification accuracy rate.

Results

We first excluded 14 participants whose datasets were incomplete (i.e. their data contained fewer than 1,661 trials); this left 498 participants. We then excluded incorrect trials and trials with response latencies faster than 200 ms or slower than 3000 ms. For the remaining correct trials, RTs more than 2.5 SDs away from each participant's mean were also treated as outliers. For the RT analyses, data trimming removed 7.3 percent (4.6 percent errors; 2.7 percent RT outliers). For all analyses, we used z-score transformed RTs, which serve to control for individual differences in processing speed (Faust et al., 1999) and to eliminate much of the variability in priming for items (Hutchison et al., 2008). Z-scores were computed separately for each participant.

Analysis 1: Reliability of Semantic Priming

Trials for each participant were first partitioned into session 1 (S1) trials, session 2 (S2) trials, odd-numbered trials, and even-numbered trials; trial number denotes the order in which the trials were presented to the participant. For each participant, we then computed the z-score priming effect (mean unrelated z-score RT—mean related z-score RT) for all trials, S1 trials, S2 trials, odd-numbered trials, and even-numbered trials, as a function of SOA and prime type. Distinguishing between S1 and S2 trials, and between odd and even-numbered trials, are prerequisites for computing test–retest and split-half reliability, respectively. Table 9.1 presents the means and standard deviations of priming effects by experimental condition and trial type (all trials, odd-numbered trials, even-numbered trials, S1 trials, S2 trials).

Subjecting the overall priming effects to a 2 (SOA) × 2 (prime type) repeated measures analysis of variance revealed main effects of SOA, $F(1, 497) = 30.00$, $p < 0.001$, $MSE = 0.012$, $\eta_p^2 = 0.06$, and Prime Type, $F(1, 497) = 70.18$, $p < 0.001$, $MSE = 0.012$, $\eta_p^2 = 0.12$; the SOA × Prime Type interaction was not significant, $F < 1$. Unsurprisingly, priming effects were larger when first associates ($M = 0.12$) were presented as targets, compared to when other associates ($M =$

TABLE 9.1 Means and standard deviations of z-score transformed priming effects as a function of experimental condition and trial type.

Lexical decision ($N = 498$)

	Overall	*Odd*	*Even*	*Session 1*	*Session 2*
Short SOA first associate					
M	0.14	0.15	0.13	0.14	0.13
SD	0.13	0.17	0.17	0.17	0.16
Short SOA other associate					
M	0.09	0.10	0.09	0.10	0.08
SD	0.11	0.16	0.15	0.15	0.15
Long SOA first associate					
M	0.11	0.12	0.10	0.12	0.10
SD	0.13	0.17	0.17	0.16	0.17
Long SOA other associate					
M	0.07	0.07	0.07	0.07	0.06
SD	0.12	0.16	0.16	0.16	0.16
All conditions					
M	0.10	0.11	0.10	0.11	0.09
SD	0.08	0.09	0.10	0.10	0.10

0.08) were used. Less expectedly, priming effects were slighter larger at the 200 ms SOA ($M = 0.11$) than at the 1200 ms SOA ($M = 0.09$). The greater priming for first-associate trials is not surprising. However, priming effects are generally larger (not smaller) at longer SOAs (Neely, 1977; but see Neely, O'Connor, & Calabrese, 2010). We will comment on this intriguing pattern in the Discussion.

It is also worth noting that the priming effects in the SPP are somewhat smaller than what one would expect, using studies such as Hutchison et al. (2008) as a reference point. As Hutchison et al. (2013) have already acknowledged, it is not entirely clear why this difference exists, but they suggest that this may be due to the fact that related trials in semantic priming experiments (e.g. Hutchison et al., 2008) typically predominantly feature very strong associates, whereas the SPP stimuli are far more diverse with respect to semantic and associative relations.

Turning to the reliability analyses, Table 9.2 presents the Pearson correlations between odd and even-numbered trials (split-half reliability), and between S1 and S2 trials (test–retest reliability), for participant-level priming effects. Like Stolz et al. (2005), we are examining correlations between responses to distinct sets of prime-target pairs, but the counterbalancing procedure ensures that the descriptive statistics of different variables are relatively similar across different sub-lists. For lexical decision, with respect to within-session reliability (reflected by split-half reliability), we observed moderate correlations (*r*s from 0.21 to 0.27) for first-associate trials, and very low correlations (*r*s from 0.07 to 0.08) for other-associate trials. Turning to between-session reliability (reflected by test–retest reliability), correlations were moderate (*r*s from 0.25 to 0.31) for first-associate trials, and very low (*r*s from 0.07 to 0.11) for other-associate trials.

Clearly, the reliabilities of semantic priming for first-associate trials are not only statistically significant, but are consistently higher than for counterpart other-associate trials (all $ps < 0.05$), whose reliabilities approach non-significance.[1] It is also reassuring that our estimates (for first-associate trials) fall broadly within

TABLE 9.2 Correlations between session 1 and session 2 participant-level priming effects, and odd and even-numbered trial participant-level priming effects.

Lexical decision ($N = 498$)

	Short SOA		*Long SOA*	
	First associate	*Other associate*	*First associate*	*Other associate*
Odd–even	0.269***	0.084†	0.207***	0.069
S1-S2	0.308***	0.107*	0.246***	0.070

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

the range of test–retest reliabilities reported by Stolz et al. (2005) for their conditions where relatedness proportion was 0.50. Specifically, for the SOAs of 200 ms, 350 ms, and 800 ms, they found test–retest correlations of 0.30, 0.43, and 0.27, respectively.

Analysis 2: Individual Differences in Semantic Priming

Consistent with Stolz et al. (2005), we found moderate-sized split-half and test–retest reliabilities for semantic priming, but only for first-associate trials. For other-associate trials, correlations were mostly non-significant, even when a relatedness proportion of 0.50 was used. These results indicate that priming is more likely to be reliable when related primes and targets are strongly associated, and reliability is present even when the SOA is very short (i.e. 200 ms). Having established that first-associate semantic priming is reliable, we next examined whether this was moderated by theoretically important individual differences. Before conducting these analyses, we excluded 17 participants who scored more than 2.5 standard deviations below the sample mean for any individual difference measure, leaving 481 participants. Table 9.3 presents the correlations between the individual difference measures, while Table 9.4 (see also Figures 9.1 and 9.2) presents the correlations between participant-level first-associate priming effects

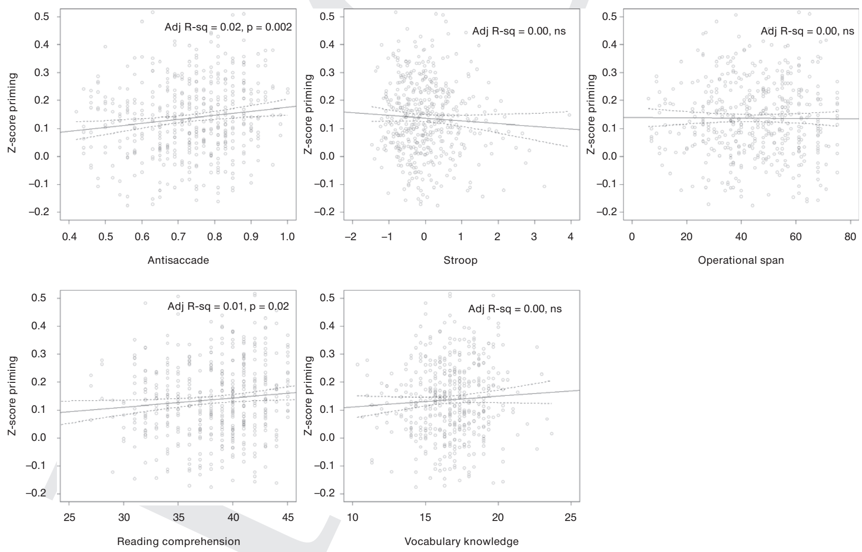

FIGURE 9.1 Scatterplots (with 95 percent confidence interval) between standardized priming effects and the five individual difference measures, when first-associate primes at an SOA of 200 ms are presented.

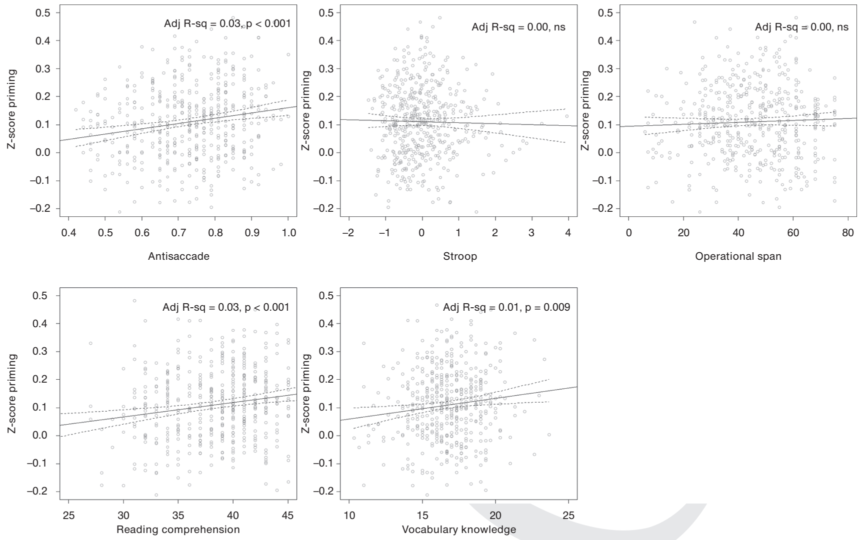

FIGURE 9.2 Scatterplots (with 95 percent confidence interval) between standardized priming effects and the five individual difference measures, when first-associate primes at an SOA of 1,200 ms are presented.

TABLE 9.3 Correlations between the individual difference measures.

	Antisaccade	*Operational span*	*Stroop*	*Reading comprehension*	*Vocabulary knowledge*
Antisaccade	–	0.273***	−0.109*	0.163***	0.234***
Operational Span		–	−0.277***	0.312***	0.374***
Stroop			–	−0.182***	−0.206***
Reading Comprehension				–	0.654***
Vocabulary Knowledge					–

*** $p < 0.001$; * $p < 0.05$

(for both short and long SOA) and attentional control measures (antisaccade, Stroop,[2] operational span), and between priming and reading ability (reading comprehension and vocabulary knowledge).[3]

In order to address the possibility that reading comprehension differences in priming are spuriously driven by differences in antisaccade performance (or vice versa), we also computed partial correlations. For short SOA priming, antisaccade

TABLE 9.4 Correlations between participant-level priming effects with attentional control and reading ability measures.

200 ms SOA ($N = 481$)

Individual difference measure	*Correlation with priming*
Attentional control	
Antisaccade	0.136**
Stroop	−0.057
Operational span	−0.007
Reading ability	
Reading comprehension	0.104*
Vocabulary knowledge	0.061

1200 ms SOA ($N = 481$)

Individual difference measure	*Correlation with priming*
Attentional control	
Antisaccade	0.181***
Stroop	−0.020
Operational span	0.042
Reading ability	
Reading comprehension	0.168***
Vocabulary knowledge	0.119**

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

performance predicted priming ($r = 0.121$, $p = 0.008$) when reading comprehension was controlled for, while the correlation between reading comprehension and priming was borderline significant ($r = 0.083$, $p = 0.068$) when antisaccade performance was controlled for. For long SOA priming, antisaccade performance predicted priming ($r = 0.159$, $p < 0.001$) when reading comprehension and vocabulary knowledge were controlled for. Reading comprehension predicted priming ($r = 0.119$, $p = 0.009$) when antisaccade performance and vocabulary knowledge were controlled for. Vocabulary knowledge did not predict priming ($r = -0.015$, *ns*) when antisaccade performance and reading comprehension were controlled.

In sum, the results are clear-cut and straightforward to summarize. Specifically, for both SOAs, two of the three measures of attentional control (Stroop and operational span) were not correlated with priming, while participants who performed better on the antisaccade task were associated with larger priming effects. Of the two reading ability measures, reading comprehension was more reliably related to priming, with reading comprehension scores correlating positively with priming effects at both SOAs. It is worth reiterating that these correlations cannot simply be attributed to scaling or to general slowing, because the participant-level priming effects were computed from z-score RTs, which controls for individual differences in processing speed (Faust et al., 1999).

Discussion

In the present study, by analyzing the large-scale behavioral data from the SPP (Hutchison et al., 2013), we examined whether semantic priming was reliable, and if so, how semantic priming was moderated by individual differences. There are a couple of noteworthy observations. First, we extend previous work on reliability (Hutchison et al., 2008; Stolz et al., 2005) by demonstrating that the reliability of semantic priming effects depends on the association strength between primes and targets. Second, we considered the impact of a broad array of individual difference measures on semantic priming, and our analyses reveal that participants with more attentional control and reading ability are associated with stronger priming.

Reliability of Semantic Priming

Stolz et al. (2005) were the first to examine the reliability of semantic priming across different levels of SOA and relatedness proportion and found that although priming effects were very robust at the level of the group, there was much less stability in the within and between-session performance of individual participants. Specifically, test–retest and split-half reliabilities were statistically significant *only* when the relatedness proportion was 0.50 (at SOAs of 200 ms, 350 ms, and 800 ms), and when a long SOA (800 ms) was paired with a high relatedness proportion of 0.75. In our study, we found moderate-sized and reliable priming with a relatedness proportion of 0.50, at both short (200 ms) and long (1200 ms) SOAs, and extended Stolz et al.'s (2005) findings by showing that significant reliability is observed only when relatively strong associates are used as related prime-target pairs.

In the account of their findings, Stolz et al. (2005) suggested that the contents of semantic memory are inherently noisy and uncoordinated, and that when priming reflects fully automatic processes (e.g. automatic spreading activation), effects can be expected to be unreliable. However, when semantic priming also reflects more strategic mechanisms such as memory recruitment and expectancy-based processes, which are sensitive to the task context, priming effects become more reliable (see Borgmann, Risko, Stolz, & Besner, 2007, which makes a similar argument for reliability in the Simon task). For example, consider the performance of a participant undergoing the primed lexical decision task on two separate occasions. Participants who are more likely to recruit the prime episode during target processing or to generate expectancies in response to the prime on the first occasion can also be expected to behave the same way on the second occasion.

This account helps provide a unified explanation for Stolz et al.'s (2005) results. That is, priming is unreliable at all SOAs when the relatedness proportion is 0.25 because there is insufficient incentive (only one in four chance the target is related to the prime) for the participant to generate potential targets or for the prime episode to be retrieved. Increasing the relatedness proportion to 0.50 drives up the likelihood of episodic prime retrieval (for short SOAs) and expectancy generation (for longer SOAs), which then yields reliable priming effects. However, one might then wonder why priming was reliable only at the longest SOA of 800 ms when the relatedness proportion was 0.75. Stolz et al. (2005) suggested that at very high relatedness proportions, participants begin to notice the prime-target relation and attempt to generate expectancies in an intentional manner. At shorter SOAs (200 ms and 350 ms), expectancy-based processes are likely to fail, because there is insufficient time for intentional generation and successful application of expectations. These failed attempts disrupt and overshadow the impact of prime retrieval, hence attenuating priming reliability.

We think our results can be nicely accommodated by a similar perspective. In the SPP, a relatedness proportion of 0.50 was consistently used, which facilitates the recruitment of the prime episode. However, why was reliability so much higher for first-associate, compared to other-associate, prime-target pairs? As mentioned earlier, prime recruitment processes are positively correlated with prime utility. That is, there is a lower probability of prime recruitment under experimental contexts where the prime is less useful, such as when the relatedness proportion is low (Bodner & Masson, 2003), when targets are clearly presented (Balota, Yap, Cortese, & Watson, 2008), and as the present results suggest, when primes are weakly associated with their targets. In general, our results provide converging support for Stolz et al.'s (2005) assertion that an increase in prime episode retrieval yields greater reliability in priming.

However, while significant test–retest and split-half correlations at the very short SOA of 200 ms (see Table 9.2) may suggest that reliability under these conditions is not mediated by expectancy, Hutchison (2007) has cautioned against relying on a rigid cutoff for conscious strategies such as expectancy generation. Indeed, it is more plausible that expectancy-based processes vary across items, participants, and practice. Consistent with this, there is mounting evidence for expectancy *even at short SOAs* for strong associates and high-AC individuals (e.g. Hutchison, 2007). We tested this by using a median split to categorize participants as low-AC or high-AC, based on their performance on the antisaccade task; reliabilities of short SOA, first-associate trials were then separately computed for the two groups. Interestingly, reliabilities were numerically higher for the high-AC group (split-half $r = 0.327$, $p < 0.001$; test–retest $r = 0.388$, $p < 0.001$) than for the low-AC group (split-half $r = 0.215$, $p < 0.001$; test–retest $r = 0.234$, $p < 0.001$); while the group difference was not significant for split-half reliability, it approached significance ($p = 0.06$) for test–retest reliability. Hence, it is possible that for high-AC

"9781138791923C09" — 2016/8/22 — 20:41 — page 219 — #17

participants, expectancy generation may have partly contributed to the reliability of short SOA priming.

Individual Differences in Semantic Priming

The literature on individual differences in semantic priming is relatively sparse, but previous work suggests that readers with higher-quality lexical representations (as assessed by vocabulary knowledge) rely more heavily on relatively automatic prospective priming mechanisms, which are not modulated by target difficulty, whereas readers with lower-quality representations show more influence of retrospective mechanisms (e.g. episodic prime retrieval or semantic matching) that increase prime reliance for more difficult targets (Yap et al., 2009). There is also evidence that individuals with more attentional control are better at strategically calibrating their reliance on expectancy-based processes in response to contextual variations in the predictive power of the prime (Hutchison, 2007).

We found that individuals who performed better on the antisaccade task also produced larger priming effects. This is consistent with a recent study by Hutchison et al. (2014), which reported greater priming for high-AC participants, but only when forward associated prime-target pairs (e.g. *atom*—*BOMB*) were presented. These results suggest that high-AC individuals are better able at prospectively generating and maintaining expectancy sets of possible related targets in response to primes, thereby increasing priming for such participants. Indeed, Heyman, Van Rensbergen, Storms, Hutchison, and De Deyne (in press) recently replicated and extended Hutchison et al. (2014) by demonstrating that imposing a high working memory load entirely eliminated forward priming, but left backward priming intact. It is not entirely clear why operation span and Stroop performance did not correlate with semantic priming effects,[4] since all three attentional capacity tasks putatively require participants to maintain information in working memory while performing an ongoing task (Hutchison, 2007). However, in a comprehensive analysis of the relationships between various working memory tasks, Shipstead et al. (2014), using structural equation modeling, reported that operational span, Stroop, and antisaccade performance do not seem to reflect a unitary construct. Instead, operational span primarily taps the primary and secondary memory components of working memory capacity (i.e. size of a person's attentional focus), whereas Stroop and antisaccade performance are more directly related to the attentional control component of working memory capacity. Notably, antisaccade performance (loading = 0.71) loaded far more highly on the AC factor than Stroop performance (loading = −0.25), and Shipstead et al. (2014) speculated that this was because the demands of the antisaccade task makes it particularly sensitive to individuals' ability to rapidly recover from attentional capture. In contrast, the Stroop task is tapping the efficiency of mechanisms that support the early inhibition of distracting information. The present analyses further reinforce the importance

"9781138791923C09" — 2016/8/22 — 20:41 — page 220 — #18

of considering the influence of AC on word recognition performance, and also suggest that researchers may want to give more weight to antisaccade performance when operationalizing AC.

In addition to attentional control, there was also evidence that more skilled lexical processors (as reflected by better performance on reading comprehension and, to a lesser extent, vocabulary knowledge measures) showed greater priming. According to the lexical quality hypothesis (Perfetti & Hart, 2001), individuals with higher quality representations should rely *less* on the strategic use of prime information for resolving targets (Yap et al., 2009), and one would therefore predict less, not more, priming for better lexical processors. However, if one assumes that higher quality representations allow *prime* words to be identified faster, thereby increasing the efficiency of prospective priming mechanisms such as automatic spreading activation and expectancy, related primes provide a greater head-start for highly skilled readers, which increases priming (Hutchison et al., 2014). This is also consistent with developmental data which indicate that as an individual's semantic network develops over the lifespan, the links between nodes in the semantic network become stronger, which in turn increases the influence of automatic priming mechanisms (Nakamura, Ohta, Okita, Ozaki, & Matsushima, 2006). Of course, the preceding *post hoc* explanation is speculative and awaits empirical verification.

Reliability of Isolated Versus Primed Lexical Decision

The results of our reliability analyses are broadly in line with Stolz et al.'s (2005) earlier findings. That is, semantic priming effects show moderate reliability ($r \approx 0.30$), but only when the relatedness proportion is sufficiently high (i.e. 0.50) and when primes and targets are strongly related. Related to this, the correlations between semantic priming effects and individual differences are similarly modest. In classical test theory, the correlation between two variables cannot exceed the square root of the product of the two variables' reliabilities (Nunnally & Bernstein, 1994). That is, even if semantic priming were correlated with a perfectly reliable measure, the correlation cannot, in principle, exceed 0.55 in absolute magnitude. In practice, when less reliable measures are examined, the correlations between attentional control and priming have ranged between 0.18 and 0.20 (see Hutchison, 2007; Hutchison et al., 2014).

These findings contrast strongly with recent work examining individual differences in *isolated* word recognition. Specifically, when one considers the stability of participants' sensitivity to lexical characteristics such as length, neighborhood characteristics, and word frequency in isolated lexical decision, correlations (*r*s between 0.38 and 0.75) are considerably higher (Yap et al., 2012). The implication here is that although spreading activation processes within the semantic network do not operate in a deterministic manner, this incoherence does not extend to the processing of a word's orthographic and phonological characteristics. Indeed, this claim meshes well with findings from Waechter et al. (2010), who showed that unlike semantic priming, repetition priming (e.g. *dog*—*DOG*) performance, which provides insights into orthographic and phonological input coding processes, was largely reliable.

Why was Priming Attenuated at a Longer SOA?

One unexpected result from the present analyses was the observation of weaker priming at a longer SOA. Priming effects at very short (< 300 ms) SOAs should primarily reflect automatic processes (but see Hutchison, 2007), whereas priming effects at longer SOAs become larger due to an added influence of controlled expectancy-based processes. Because the strategic processes which underlie the generation of potential candidates from the prime require time to develop, they should therefore be most influential at long SOAs (Neely, 1977). We considered a couple of possible explanations for these results.

First, as mentioned before, in typical semantic priming experiments, strongly associated prime-target pairs are overrepresented as stimuli. In the SPP, the large number of stimuli featured a full range of associative strength and feature overlap, and this perhaps made it more difficult for participants to generate and maintain expectancies. However, as reported in the Results section (see Analysis 1), prime type and SOA produced additive effects on semantic priming, that is, SOA decreased priming to an equivalent extent for first-associate and other-associate pairs. If prime-target relatedness did matter, then one would predict a larger effect of SOA for other-associate trials. Second, the participants in the SPP were exposed to an atypically large number of trials (over 800 trials per session) over two sessions, and it is possible that this caused participants to stop maintaining expectancies over time. Additional analyses provide some evidence for this. Specifically, the effect of session was significant, $F(1, 497) = 9.27$, $p = 0.002$, $MSE = 0.022$, $\eta_p^2 = 0.02$, with less priming at session 2 ($M = 0.09$) than at session 1 ($M = 0.11$).

Limitations and Future Directions

In the present study, we examined trial-level data from the SPP, an online behavioral repository of over 800,000 primed lexical decision trials from over 500 participants. In line with previous studies, we found evidence that semantic priming effects are moderately reliable, but mainly under conditions which foster reliance on episodic prime retrieval or the strategic generation of related targets. We also observed greater priming for participants with superior attentional control and higher quality lexical representations, suggesting that such individuals are better able to take advantage of prime information to drive prospective priming mechanisms. From a more methodological perspective, this work strongly underscores the utility of the megastudy approach for investigating semantic priming phenomena, and we look forward to seeing more interesting questions being answered by the SPP.

Of course, there are a number of limitations associated with the present work. First, due to the archival nature of the SPP dataset, we were only able to explore the influence of the individual differences measures available. Researchers (Perfetti, 1992) have proposed that the most faithful index of lexical quality is spelling performance, because accurate spelling requires precise lexical representations; performance on other tasks (e.g. vocabulary knowledge, reading comprehension) can be achieved on the basis of partial lexical information, which can be compensated for by context (Andrews, 2012). In future work on individual differences in semantic priming, researchers may want to explore the role of spelling performance, which is presumably a more direct measure of lexical quality.

Second, we did not distinguish between forward, backward, and symmetrically associated prime-target pairs in our analyses. Recent work (Heyman et al., in press; Hutchison et al., 2014; Thomas, Neely, & O'Connor, 2012) makes it clear that prime direction can modulate the influence of automatic and controlled priming processes. Future investigations can consider the interplay between attentional control, lexical quality, and priming mechanisms in a finer-grained manner by taking prime direction into account. Third, while we agree with Stolz et al. (2005) that the present empirical evidence is most consistent with a noisy and uncoordinated semantic system, it is difficult to entirely eliminate the effects of controlled processing, even with a short SOA and low relatedness proportion. To provide less equivocal support for the instability of automatic priming, it will be useful to conduct similar reliability analyses using masked priming paradigms that minimize the contaminating influence of strategy (Forster, 1998). Finally, despite the scope of the SPP, one could plausibly argue that a sample of college students, who are to a large extent selected for their reading ability and attentional control, will show a restricted range of these measures compared to the general population. Hence, it is possible that the true relationships between reading ability, attentional control, and priming may be even larger than we have reported here.

Notes

1 Given the weaker priming observed for other-associate, compared to first-associate, trials, one might wonder if reliability is lower in this condition because of decreased variability in priming across participants. As suggested by a reviewer, this possibility is ruled out by our data (see Table 9.1), which reveal comparable variability in priming for first and other-associate trials.

2 Stroop performance was computed by averaging standardized Stroop effects in RTs and accuracy rates.

3 Vocabulary knowledge was computed by averaging scores on the synonym, antonym, and analogy tests.
4 We also explored this by examining the correlations between priming effects and performance on the three tasks in previous studies (Hutchison, 2007; Hutchison et al., 2014). While antisaccade performance was overall the strongest predictor of priming, all three measures correlated with priming in at least one of the experiments. Of course, the priming effects in these studies were based on a smaller number of observations (20 related and 20 unrelated trials).

References

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703–719.
Andrews, S. (2012). Individual differences in skilled visual word recognition and reading: The role of lexical quality. In J. S. Adelman (Ed.), *Visual word recognition volume 2: Meaning and context, individuals, and development* (pp. 151–172). Hove: Psychology Press.
Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General, 139*, 299–318.
Balota, D. A., & Yap, M. J. (2006). Attentional control and flexible lexical processing: Explorations of the magic moment of word recognition. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 229–258). New York: Psychology Press.
Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language, 59*, 495–523.
Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459.
Balota, D. A., Yap, M. J., Hutchison, K.A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In James S. Adelman (Ed.), *Visual word recognition Volume 1: Models and methods, orthography and phonology* (pp. 90–115). Hove: Psychology Press.
Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8*, 493–512.
Bodner, G. E., & Masson, M. E. J. (1997). Masked repetition priming of words and non-words: Evidence for a nonlexical basis for priming. *Journal of Memory and Language, 37*, 268–293.
Bodner, G. E., & Masson, M. E. J. (2003). Beyond spreading activation: An influence of relatedness proportion on masked semantic priming. *Psychonomic Bulletin and Review, 10*, 645–652.
Borgmann, K. W. U., Risko, E. F., Stolz, J. A., & Besner, D. A. (2007). Simons says: Reliability and the role of working memory and attentional control in the Simon Task. *Psychonomic Bulletin and Review, 14*, 313–319.
Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory and Cognition, 7*, 68–76.

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

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

Forster, K. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research, 27*, 203–233.

Goodwin, C. J. (2009). *Research in psychology: Methods and design* (6^{th} edn.). Hoboken, NJ: Wiley.

Heyman, T., Van Rensbergen, B. V., Storms, G., Hutchison, K. A., & De Deyne, S. (2015). The influence of working memory load on semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*, *41*, 911–920.

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 645–662.

Hutchison, K. A., Balota, D. A., Cortese, M., & Watson, J. M. (2008). Predicting semantic priming at the item-level. *The Quarterly Journal of Experimental Psychology, 61*, 1036–1066.

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C-S., . . . Buchanan, E. (2013). The Semantic Priming Project. *Behavior Research Methods, 45*, 1099–1114.

Hutchison, K. A., Heap, S. J., Neely, J. H., & Thomas, M. A. (2014). Attentional control and asymmetric associative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 844–856.

Kinoshita, S., Forster, K. I., & Mozer, M. C. (2008). Unconscious cognition isn't that smart: Modulation of masked repetition priming effect in the word naming task. *Cognition, 107*, 623–649.

Kinoshita, S., Mozer, M. C., & Forster, K. I. (2011). Dynamic adaptation to history of trial difficulty explains the effect of congruency proportion on masked priming. *Journal of Experimental Psychology: General, 140*, 622–636.

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General, 122*, 316–330.

Lowe, C., & Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia, 36*, 915–923.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Hove, UK: Psychology Press.

Matthews, G., & Harley, T. A. (1993) Effects of extraversion and self-report arousal on semantic priming: A connectionist approach. *Journal of Personality and Social Psychology, 65*, 735–756.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

Morgan, C. J. A., Bedford, N. J., & Rossell, S. L. (2006). Evidence of semantic disorganization using semantic priming in individuals with high schizotypy. *Schizophrenia Research, 84*, 272–280.

Nakamura, E., Ohta, K., Okita, Y., Ozaki, J., & Matsushima, E. (2006). Increased inhibition and decreased facilitation effect during a lexical decision task in children. *Psychiatry and Clinical Neurosciences, 60*, 232–239.

Nation, K., & 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. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 236–264). Hillsdale, NJ: Erlbaum.

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1003–1019.

Neely, J. H., O'Connor, P. A., & Calabrese, G. (2010). Fast trial pacing in a lexical decision task reveals a decay of automatic semantic activation. *Acta Psychologica, 133*, 127–136.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36, 402–407.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd edn.). New York: McGraw-Hill.

Payne, B. K. (2005) Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology, 89*, 488–503.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A, & Hart, L. (2001). The lexical bases of comprehension skill. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67–86). Washington, DC: American Psychological Association.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107*, 786–823.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. L. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72*, 116–141.

Stolz, J. A., Besner, D., & Carr, T. H. (2005). Implications of measures of reliability for theories of priming: Activity in semantic memory is inherently noisy and uncoordinated. *Visual Cognition, 12*, 284–336.

Thomas, M. A., Neely, J. H., & O'Connor, P. (2012). When word identification gets tough, retrospective semantic processing comes to the rescue. *Journal of Memory and Language, 66*, 623–643.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505.

Waechter, S., Stolz, J. A., & Besner, D. (2010). Visual word recognition: On the reliability of repetition priming. *Visual Cognition, 18*, 537–558.

Whittlesea, B. W. A., & Jacoby, L. L. (1990). Interaction of prime repetition with visual degradation: Is priming a retrieval phenomenon? *Journal of Memory and Language, 29*, 546–565.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock Johnson III tests of cognitive abilities*. Rolling Meadows, IL: Riverside Publishing.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognitions: Insights from the English lexicon project. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 53–79.

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to non-words in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*, *41*, 597–613.

Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency: The role of lexical integrity. *Journal of Memory and Language, 61*, 303–325.