The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

The role of lexical variables in the visual recognition of Chinese characters: A megastudy analysis

Wei Ping Sze[a], Melvin J. Yap[a] & Susan J. Rickard Liow[a]
[a] Department of Psychology, National University of Singapore, Singapore
Published online: 27 Jan 2015.

To cite this article: Wei Ping Sze, Melvin J. Yap & Susan J. Rickard Liow (2015): The role of lexical variables in the visual recognition of Chinese characters: A megastudy analysis, The Quarterly Journal of Experimental Psychology, DOI: 10.1080/17470218.2014.985234

To link to this article: http://dx.doi.org/10.1080/17470218.2014.985234

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2015
http://dx.doi.org/10.1080/17470218.2014.985234

The role of lexical variables in the visual recognition of Chinese characters: A megastudy analysis

Wei Ping Sze, Melvin J. Yap, and Susan J. Rickard Liow

Department of Psychology, National University of Singapore, Singapore

Logographic Chinese orthography partially represents both phonology and semantics. By capturing the online processing of a large pool of Chinese characters, we were able to examine the relative salience of specific lexical variables when this nonalphabetic script is read. Using a sample of native mainland Chinese speakers ($N = 35$), lexical decision latencies for 1560 single characters were collated into a database, before the effects of a comprehensive range of variables were explored. Hierarchical regression analyses determined the unique item-level variance explained by orthographic (frequency, stroke count), semantic (age of learning, imageability, number of meanings), and phonological (consistency, phonological frequency) factors. Orthographic and semantic variables, respectively, accounted for more collective variance than the phonological variables. Significant main effects were further observed for the individual orthographic and semantic predictors. These results are consistent with the idea that skilled readers tend to rely on orthographic and semantic information when processing visually presented characters. This megastudy approach marks an important extension to existing work on Chinese character recognition, which hitherto has relied on factorial designs. Collectively, the findings reported here represent a useful set of empirical constraints for future computational models of character recognition.

Keywords: Mandarin; Visual word recognition; Lexical decision; Nonalphabetic; Logograph.

Valuable insights into the processing of alphabetic languages have been gained by examining the influence of different variables using the megastudy approach. Instead of researchers selecting stimuli based on a limited set of criteria, this methodology allows the language to define the stimuli (see Balota, Yap, Hutchison, & Cortese, 2012, for a review). Although Chinese, a morphosyllabic language with logographic orthography, is used by over a billion people (Li & Thompson, 2009), research directly examining how various lexical variables influence Chinese character recognition has been largely confined to factorial experiments. To our knowledge, only Y. Liu, Shu, and Li (2007) have examined a range of variables at the same time, and their study involved the naming paradigm. Importantly, very few researchers have employed the lexical decision task during their work on Chinese character recognition. This is surprising, given that lexical decision is probably the most popular paradigm used for exploring the processing of alphabetic orthographies. The results of a recent search, using the Web of Science database, revealed that since the 1980s, there have been only 35 lexical decision studies carried out using single characters, and all of these involved small-scale factorial analyses. The Chinese Lexicon Project (Sze, Rickard Liow, & Yap, 2014) was

Correspondence should be addressed to Wei Ping Sze, Department of Psychology, National University of Singapore, Block AS4, #02-07, 9 Arts Link, Singapore 117570. E-mail: weipingsze@gmail.com

specifically developed to address this gap. In this article, we report the item-level performance of orthographic, semantic, and phonological predictors on visual lexical decision latencies for 1560 single Chinese characters. Several outstanding theoretical issues related to a range of lexical variables and the interactions between them are also examined.

There are at least three reasons why the item-level analyses in this study are important and timely. First, the findings should help contribute to a deeper understanding of the processes involved in reading a nonalphabetic orthography. These processes are likely to be different from those employed for alphabetic orthographies, where the emphasis is on the mapping between spelling and sound (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Although there is a growing appreciation of the need to understand how nonalphabetic writing systems are decoded before making language-general claims (Perfetti, Zhang, & Berent, 1992; Seidenberg, 1985), an Anglocentric view of word recognition continues to dominate research (see Share, 2008, on the problems associated with the overreliance on research in English as a guide to reading processes in other languages). In order to begin exploring the relative impact of different psycholinguistic variables on the processing of Chinese characters, it is crucial to understand the core elements making up this unusual writing system. In Chinese, single characters are the most basic lexical unit that can be used to construct "words". For example, "理" (/li3/, meaning "reason") and "智" (/zhi1/, meaning "intelligence") concatenate to form the double-character word "理智" (which means "sanity"). As can be seen, each character simultaneously encodes, albeit indirectly, both syllable phonology and semantics.

Linguistically, this relationship between sound and meaning is often described as "morphosyllabic"—that is, the single character represents a unit of meaning (morpheme), as well as a single phonetic syllable (Hoosain, 1991). About 81% of single characters are composed of a semantic and a phonetic component, and these are commonly referred to as semantic and phonetic radicals, respectively (e.g., Y. Zhou, 2003, p. 84). This typical semantic-plus-phonetic single-character structure reinforces the notion that there are close ties between orthography, semantics, and phonology in the Chinese writing system. However, the relative impact of each element within this orthography–semantics–phonology trinity on character processing remains unclear. There are two main schools of thought: the traditional semantic account and the phonological account (Seidenberg, 1985), both of which are reviewed later.

The second reason for conducting a large-scale analysis is because the megastudy approach has the potential to validate (or invalidate) the existing findings from factorial experiments that dominate the field currently. Unlike psycholinguistic research in other languages, say, English or French, work on Chinese character processing has been founded almost entirely on factorial designs. With the exception of Y. Liu et al.'s (2007) naming study and Sze et al.'s (2014) lexical decision research, all experiments thus far have been cast in the *x* by *y* factorial grid with a limited sample of stimuli. These factorial experiments, typically analysed using analysis of variance (ANOVA), have produced rich insights, but a megastudy approach will provide convergent validity for current models of cognitive processing.

A brief summary of the potential drawbacks related to factorial experiments, specific to Chinese psycholinguistics, is worthwhile (discussion on the limitations of factorial designs in psycholinguistics is not new; see Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004, for a comprehensive overview). Factorial studies assume that all stimuli are either equally matched or manipulated in terms of important lexical characteristics, so that they can be fitted into the respective experimental cells. This is sometimes not logistically possible when compiling materials for experimental work. Careful matching produces its own constraints because the small pool of selected stimuli may not be representative of the average reader's lexicon. Hue (1992) for instance, highlighted Fang, Horng, and Tzeng's (1986) failure to fully account for character frequency. The difficulty in matching stimuli across conditions is an issue that is revisited later.

Factorial designs also necessitate the categorization of continuous variables into two or more bands—for example, high/medium/low. This is a problem because many lexical variables have a continuous scale of measurement. Consistency, for example, is a continuous variable, and so is printed frequency. As Cohen (1983) pointed out, turning continuous variables into discrete categories results in a loss of information. Artificially imposing categories could also distort the effects, yielding misleading results (see MacCallum, Zhang, Preacher, & Rucker, 2002, for related simulations).

Problems with matching and artificial dichotomies could explain why findings for Chinese character processing have sometimes been contradictory (see I.-M. Liu, 2003, for a brief review). Factorial studies are valuable because they allow researchers to target a specific variable of interest on a manageable scale, but ideally they should be complemented by other designs. The megastudy approach described in this paper is one alternative with several benefits. In terms of scope, large-size databases are likely to be more representative of the typical reader's lexicon of 4000 characters (estimated to be sufficient for recognizing 99.9% of characters encountered; Yin & Rohsenow, 1994, p. 69) because they are less susceptible to sampling bias. In addition, having a large dataset allows experimenters to expand their repertoire of statistical techniques beyond factorial analyses. Regression techniques, for example, allow researchers to avoid the design pitfalls of weak stimuli matching and rigorously evaluate each predictor, while holding the other important predictors constant. A megastudy approach also frees the lexical items from the constraints imposed by experimental manipulation. It returns predictive power back to the language and allows the most salient lexical characteristics to influence the outcome (also see Balota et al., 2012, for a recent review on megastudies of other languages; Keuleers, Lacey, Rastle, & Brysbaert, 2012, for further examples of the advantages).

The third reason why a large-scale item-level analysis is important and timely is because it provides a summary of the key lexical variables involved in the visual recognition of Chinese characters. Weighting a list of possible lexical phenomena is crucial because the separate roles performed by each predictor (or predictor set) can be evaluated, while holding other variables constant. These results are critical for benchmarking the lexical effects deemed salient for skilled readers. An ideal cognitive or computational model should be capable of explaining and/or simulating the salient lexical effects identified by online behavioural testing. This approach has been adopted by researchers when validating the dual-route cascaded (Coltheart et al., 2001), the Plaut, Seidenberg, McClelland, and Patterson (PSMP) (Plaut, McClelland, Seidenberg, & Patterson, 1996), and the connectionist dual process (CDP+) models (Perry, Ziegler, & Zorzi, 2007) for the reading of English words. In all three instances, the researchers sought to demonstrate the number of successful lexical effects that their respective models could replicate as way of substantiating validity (also see Coltheart, 2012).

Table 1 provides an overview of the major lexical effects previously found in small-scale factorial experiments using Chinese single characters.

In addition to the major effects listed in Table 1, three interactions—frequency by stroke count , frequency by number of meanings, and frequency by consistency—are examined. The theoretical implications of these three interactions are discussed later. In what follows, we address two issues specific to Chinese psycholinguistics that arose when the lexical variables were previously examined.

Issue 1: Semantics versus phonology debate

There is continued interest in comparing the relative importance of the orthographic, semantic, and phonological lexical constituents in Chinese character reading. The salience of orthographic predictors, such as the number of strokes and the orthographic print-frequency, is widely accepted (see Honorof & Feldman, 2006, on the number of strokes; Cai & Brysbaert, 2010, on print-frequency). For this reason, these two predictors are typically treated as control variables in experimental studies. Thus, the main debate within the literature

Table 1. *Notable lexical effects identified by single-character studies in Chinese*

Lexical Variables	*Expected finding and the representative studies*
Orthographic variables	
Character frequency	
Refers to the number of times a character is encountered.	Expected finding: Characters with higher frequencies are processed faster. Effect is very robust.
	Naming: Yu and Cao (1992); Y. Liu et al. (2007)
It indicates print exposure and is routinely applied as a control variable.	Lexical decision: Chung, Cheng, and Leong (1988, Experiment 1)
	Lexical naming: I.-M. Liu, Wu, and Chou (1996)
Number of strokes	
Refers to the numerical count of the lines taken to form a character. A stroke is a line (straight or curved; horizontal, vertical, or diagonal) completed each time the writing instrument is lifted from the written surface.	Expected finding: Characters with more strokes are processed more slowly. Effect is very robust. Naming: Yu and Cao (1992); Shen and Zhu (1994); Y. Liu et al. (2007)
Stroke count provides a measure of a character's orthovisual complexity and is routinely applied as a control variable.	Lexical decision: Peng and Wang (1997, Experiment 2); Chiang (2003, Experiment 2)
	Yes/no judgement task (fifth graders had to verify whether the two characters on display were identical): Chiang (2003, Experiment 1)
Phonological variables	
Regularity	
A categorical variable that refers to whether the pronunciation of character's phonetic stem is identical to the pronunciation of the full character. For example, "町" /ding1/ is regular, because its phonetic stem "丁" /ding1/ has the same syllable phonology (tone is typically excluded).	Expected finding: Regular characters should be processed faster. Naming: Seidenberg (1985); Fang et al. (1986)
Consistency	
A continuous variable that indicates how systematic a character's phonetic stem informs the reader of the full character's pronunciation (see "Method").	Expected finding: Consistent characters should be processed faster than inconsistent ones. Naming: Fang et al. (1986); Hue (1992); Yang, McCandliss, Shu, and Zevin (2009)
Homophone density	
Homophones are orthographically distinct single characters that happen to sound alike in terms of syllable and tone phonology, e.g., 画 /hua4/ and 桦 /hua4/.	Expected finding: Characters with more homophone mates should be processed faster. Naming: Ziegler, Tan, Perry, and Montant (Experiment 1, 2000); H.-C. Chen, Vaid, and Wu (Experiment 2, 2009)
Homophone density thus refers to the number of homophone mates a single character has. For example, assuming there are only two characters (伐 and 乏) in the entire lexicon pronounced as /fa2/, the homophone density for each of these two characters would be "2".	Lexical decision: Ziegler et al. (Experiment 1, 2000); H.-C. Chen et al. (Experiment 1, 2009)

(*Continued overleaf*)

Table 1. Continued.

Lexical Variables	*Expected finding and the representative studies*
	Discrepancies: Naming and lexical decision: Homophone density effect observed only for low-frequency characters in H.-C. Chen et al. (2009). Naming: Effect not observed in Y. Liu et al. (2007).
Phonological frequency	
Refers to the cumulative frequency for all homophone mates with the same syllable and tone phonology.	Expected finding: Characters with higher phonological frequencies should be processed faster. Lexical decision: Ziegler et al. (Experiment 2, 2000)
Returning to the /fa2/ illustration above, if 伐 has a print frequency of 340 and 乏, 888; the phonological frequency for each character will be 1228.	Discrepancies: Naming: H.-C. Chen et al. (2009) and Y. Liu et al. (2007) observed inhibitory effects for phonological frequency, not facilitatory effects as reported by Ziegler et al. (2000). Lexical decision: Effect not observed in H.-C. Chen et al. (2009)
Semantic variables	
Number of meanings	
Refers to how many distinct conceptual representations a character has, i.e., how polysemous a character is. If a character or word has multiple meanings (i.e., polysemous), it should receive a faster response because having multiple semantic entries benefits memory retrieval (see Lichacz, Herdman, Lefevre, & Baird, 1999; Rubenstein, Garfield, & Millikan, 1970, for research in English lexical decision and naming, respectively).	Expected finding: Characters with more meanings should be processed faster. Naming, lexical decision, and lexical naming: Peng, Deng, and Chen (2003) Discrepancy: Naming: Effect not observed in Y. Liu et al. (2007), but note in-text discussion later
Imageability	
Refers to a character's subjectively rated capacity to stimulate a mental referent. The referent can be either a picture or sensory experience, such as a smell or sound.	Expected finding: Highly imageable characters should be processed faster. Naming: Y. Liu et al. (2007)
Age of learning (AoL) AoL references the time frame in which a character was first introduced to a person through formal instruction. Conceptually, it is similar to the age of acquisition (AoA), except that AoA is subjective (i.e., participants are asked to estimate the age at which they first learnt a particular character, see Gilhooly & Logie, 1980) but AoL is objective and obtained from grade-level records. AoL is classified as a semantic variable in this paper[a], based on research that had identified AoA effects beyond the effects of cumulative frequency (e.g., Ghyselinck, Lewis, & Brysbaert, 2004, in Dutch). Results in favour of supporting AoA as a semantic variable were also obtained in Chinese psycholinguistics (B. Chen, You, & Zhou, 2007).	Expected finding: Characters acquired earlier (lower AoL values) should be processed faster. Naming: Y. Liu et al. (2007); B. G. Chen, Zhou, Dunlap, and Perfetti (Experiment 1, 2007) Lexical decision: B. Chen, Dent, You, and Wu (Experiment 3, 2009) Semantic categorization: B. G. Chen, Zhou, Dunlap, and Perfetti (Experiment 2, 2007); B. Chen, You, & Zhou (Experiment 2, 2007)

[a]The pattern of our results remained unchanged, even when AoL was treated as a control variable, instead of a semantic variable. When number of meanings and imageability were entered last (Step 3) in a separate set of hierarchical regression (with AoL entered in Step 1 as control variable, alongside character frequency and number of strokes), these two semantic variables still managed to explain significant amount of variance ($\Delta R^2 = .003$, $p = .02$) beyond the phonological variables (which were entered in Step 2).

is on the relative importance of semantic and phonological factors. Very briefly, we highlight the main points of each perspective below.

The semantic account in single-character recognition

According to this account, lexical access in Chinese character recognition is necessarily mediated by semantics for three reasons. First, given the prevalence of homophony in spoken Chinese, it is unclear how access to a character's lexical representation could be accomplished reliably through its phonology. The extent of homophony depends on the size of the corpus. Figures drawn from the *Dictionary of Chinese Character Information* (Science Publishers, 1998) estimate that approximately 21 characters share one pronunciation, whilst values from the *Dictionary of Modern Chinese Frequency* (Language Teaching and Research Institute of Beijing Language and Culture University, 1986) work out to roughly 11 characters per pronunciation (tones not included in either count). Thus, the argument favouring semantic-mediation is that identifying a character, like 积 (/ji1/, meaning "accumulate"), using its phonology would be ineffective and error prone, given that 37 other characters are pronounced in the same way. Secondly, the character's morphosyllabic nature underpins the close relationship between orthography and semantics. Citing pictographs, like 山 (/shan1/, meaning "mountain") as examples, some researchers have claimed that because Chinese characters are in fact graphic depictions of meanings, they evoke mental referents readily (e.g., W. S.-Y. Wang, 1973). Most importantly, the presence of semantic radicals within characters has led to the claim that lexical access is likely to be meaning based. Take for example, the character "踩" /cai3/, which means "step". It contains the semantic radical "⻊", which denotes "foot", so it appears that the semantic information is embedded within the character's orthography.

In terms of neuropsychological evidence for the semantic account, Han, Bi, Shu, and Weekes (2005) described a dyslexic patient who was able to pronounce a Chinese word accurately, provided that he could retrieve partial semantic information about the word. This patient's behavioural data suggest that Chinese reading is semantically mediated. Electrophysiological evidence for direct semantic access from orthography in single-character processing was also reported by K. Wang (2011), when event-related potentials were measured from 18 native-speakers performing a Stroop task (also see K. Wang, Mecklinger, Hofmann, & Weng, 2010, which produced similar evidence using a different set of stimuli). Combined with the linguistic arguments, it seems that native users of Chinese are likely to rely heavily on semantics when processing their written language.

The phonological account in single-character recognition

Empirical findings emerged to challenge the semantic account (e.g., Perfetti & Tan, 1998; Ziegler et al., 2000). Researchers argue that pictographic characters are rare and constitute only 4% of the character corpus (Sun, 2006), and that semantic radicals are not always helpful in representing the character's meaning. For example, 字 (/zi4/, meaning "word") has a semantic radical "宀", which is associated with "houses". Importantly, experimental work has been reported to substantiate the phonological account, including the time course for phonological activation in priming studies. Using a primed naming paradigm, Perfetti and Tan (1998) showed that phonological activation temporally precedes semantic activation. This finding appears to indicate that the phonological code is important and might even constrain semantic access (but see H.-C. Chen & Shu, 2001; X. Zhou & Marslen-Wilson, 2000, for counterevidence).

More pertinent to the present paper is the line of research espousing the role of a particular phonological variable. Typically, researchers have taken phonological variables that are well studied in the literature on English word recognition and extrapolated them to Chinese lexical processing. Seidenberg (1985), in one of the earliest papers of this type, extended the concept of regularity (Coltheart, 1978) to Chinese characters (see Table 1 on how regularity was redefined and applied to Chinese characters) and found that his participants named low-frequency regular characters significantly faster than low-frequency irregular characters. No difference was observed for the high-frequency items. He concluded that there is greater similarity between the processing of Chinese and English than formerly assumed, because regularity effects are also confined to the reading of low-frequency English words by skilled adults.

In the same vein, Fang et al. (1986) borrowed Glushko's (1979) concept of consistency in English. Through the careful manipulation of consistency within regular characters, Fang et al. (1986) discovered that participants named regular-consistent characters significantly faster than regular-inconsistent ones. Consistency might therefore have been the driving force behind Seidenberg's regularity effect (see also Hue, 1992). Similarly, Ziegler et al. (2000) tested and obtained significant phonological effects of homophone density and phonological frequency using Chinese characters.

To summarize, evidence in support of phonological recoding of Chinese characters has been marshalled across a range of phonological variables. Several of the findings have been taken to support the universal phonological principle, which is a strong form of the phonological account (e.g., Seidenberg, 1985; Ziegler et al., 2000). The universal phonological principle holds that the script–sound relation is more reliable than the script–meaning relationship (Perfetti & Tan, 1998). If strong evidence for phonological effects were to be found for reading a relatively sound-opaque writing system, such as Chinese characters, it would be reasonable to suggest that phonology plays a fundamental role in word recognition and is universal across orthographies (Perfetti & Dunlap, 2008).

Given that comparisons weighing the relative importance of the semantic and phonological lexical constituents in Chinese characters are still ongoing (e.g., Williams & Bever, 2010), a major contribution of the data from this megastudy will be to adjudicate the roles played by the semantic and phonological constituents within a character. Critically, the empirical evidence substantiating the first group of phonological variables (i.e., consistency and regularity, Table 1) has been grounded in the naming paradigm. Naming necessarily involves the explicit computation of phonology before production; it is unclear whether either consistency or regularity effects will be observed tasks, such as lexical decision, where phonological computation appears optional. Nevertheless, it has also been argued that, "lexical decision [is] the more appropriate task [than naming] to reveal early phonology, as it avoids the articulatory processes that mandate phonological processing" (see Halderman, Ashby, & Perfetti, 2012, p. 210).

Issue 2: Clarification of mixed findings

There are discrepant findings in extant literature as shown in Table 1. The present megastudy analysis could be instrumental in clarifying these inconsistencies. Most notably, Ziegler et al. (2000) found facilitatory effects for homophone density across both lexical decision and speeded naming. This was replicated by H.-C. Chen, Vaid, and Wu (2009), with a better-controlled set of stimuli that accounted for componential characteristics. Again in both character lexical decision and naming, the facilitatory effect of homophony was found, although it was restricted to low-frequency characters only. However, facilitation was not observed in a third paper, Y. Liu et al.'s (2007) mega-naming study. After controlling for other variables (e.g., number of strokes, regularity, number of meanings, imageability), they showed that homophone density failed to predict naming performance. This raises the possibility that confounds in factorial designs, such as the inadequate control for all relevant predictors or limited representativeness, could be producing conflicting results in a number of published experiments on Chinese character processing.

Evidence for an effect of phonological frequency is also mixed. There are three papers investigating this variable, two of which were based on lexical decision, and their findings contradict. Having controlled for print frequency, Ziegler et al. (2000) discovered a facilitatory effect for characters with high phonological frequency. They suggested that a character's homophony might confer processing advantages and took this as evidence for obligatory phonological activation when processing logographs. However, H.-C. Chen, et al. (2009), failed to replicate and reported null effects, so there is no consensus yet on the behaviour of phonological frequency in lexical decision. Further investigation is warranted, especially given how the effects were presented as strong evidence for the phonological account of character processing (see Ziegler et al., 2000).

A third discrepancy pertains to the number of meanings. Evidence for polysemy effects in Chinese single characters seems to rest on only two papers.[1] Peng et al. (2003) investigated the effects of this semantic variable across three paradigms (lexical decision, speeded naming, and lexical naming tasks) and found a facilitatory effect in all three tasks. However, in the meganaming study by Y. Liu et al. (2007), number of meanings was operationalized categorically, and the authors failed to detect an effect of polysemy. It is unclear, at this stage, what impact differences in methodology had on Liu et al.'s regression results, but converting a continuous variable into discrete categories reduces its variance and could have diminished the predictive power of polysemy (see Cohen, 1983, or MacCallum et al., 2002, for a related discussion). To distinguish the effects of polysemy, the raw count (rather than categories) is used in our analyses.

One further discrepancy in the literature on Chinese character processing pertains to the reported Frequency × Number of strokes interaction. This interaction has been cited as key evidence for the horse-race model of processing (Taft & Zhu, 1997, p. 763), which postulates how characters could either be decomposed into subcomponents or be treated holistically during processing (to be elaborated later). Data from the present megastudy permitted an examination of this important interaction.

THE PRESENT STUDY

Given the limitations of factorial designs, we employed the megastudy approach and a series of regression analyses on lexical decision to establish the relative influence of seven theoretically important predictors on single-character recognition:

> *Orthographic variables*—the number of strokes and print frequency.
> *Semantic variables*—age of learning, imageability, and the number of meanings.
> *Phonological variables*—consistency and phonological frequency. (Note that only two phonological variables were selected to represent the four main phonological variables listed in Table 1. Consistency was chosen over regularity and phonological frequency over homophone density to avoid multicollinearity.)

As phonological factors would be evaluated alongside the orthographic and semantic constituents, part of the analysis examines the robustness of the phonological effects and, by extension, evaluates the universal phonological principle. For the purpose of this examination, two specific predictions were tested—if phonology genuinely underpins Chinese character processing (e.g., Perfetti, Liu, & Tan, 2005; Ziegler et al., 2000), then (a) phonological predictors, in combination, should contribute additional unique variance over and beyond that of the orthographic and semantic factors; (b) at the individual level, each phonological predictor should also wield considerable predictive power. Given the firm position of the universal phonological principle, these hypothesized effects should endure, even when the task at hand does not explicitly require phonology to be computed.

Aside from the main effects of individual predictors, theoretically motivated interaction effects are also examined. These include the frequency by number of strokes, frequency by number of meanings, and frequency by consistency interactions. While the first two interactions are supported by previous research using speeded naming and lexical decision, it should be

[1]There is a more substantial body of research on this for double-character Chinese words. In general, the papers report that participants responded to polysemous words faster (e.g., in lexical decision tasks: Experiment 1 in Chen & Peng, 2001; Experiments 1 and 2 in Liu & Peng, 2005).

emphasized that to our knowledge, the frequency by consistency interaction has only been examined in speeded naming, but not in lexical decision. All three interactions are described in greater depth later. Understanding these interactions help researchers gain a more complete picture of visual character recognition, and together with the main effects of the seven variables, the megastudy provides well-defined empirical constraints for evaluating computational models in the future[2] (Balota & Spieler, 1998; Spieler & Balota, 1997; but also see Seidenberg & Plaut, 1998). To date, there are no computational models developed specifically to account for lexical decision data with Chinese single characters. Perfetti et al.'s (2005) lexical constituency model is a primed naming model, while Taft, Zhu, and Peng's (1999) multilevel activation framework is a theoretical model that is not implemented computationally. However, as these two models are popularly cited in extant literature, the associated implications of our findings will be discussed.

Method

Behavioural data from the Chinese Lexicon Project (Sze et al., 2014) were used in the present analysis. The Chinese Lexicon Project is a repository of behavioural information (accuracy and response latencies) for 2500 commonly used simplified Chinese single characters obtained from 35 native-speakers of Mandarin from mainland China. Over three sessions on separate days, these participants completed 5000 randomized lexical decision trials (2500 genuine characters, 2500 pseudocharacters created by switching semantic radicals (部首, bu4 shou3) within each pair, taking care that the radical position was maintained). Forty practice trials were also included at the start of each session. Figure 1 depicts the framework of each trial. Readers are advised to refer to Sze et al. (2014) for more details.

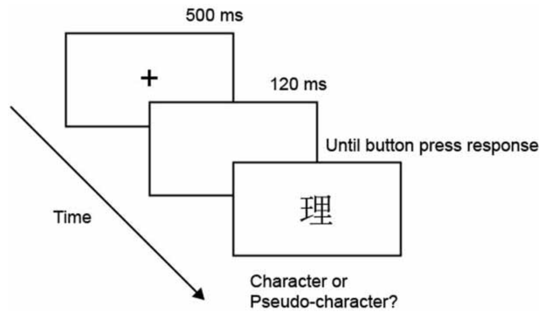

Figure 1. *Layout of the lexical decision trial.*

Results and discussion

In what follows, only response latencies from the correct character trials ("yes" responses) were analysed. On average, accuracy rates for the character and pseudocharacter trials were high: .952 ($SD = .016$) and .951 ($SD = .019$), respectively. In addition, responses faster than 200 ms or slower than 3000 ms were identified as extreme and were removed. Having addressed the extreme scores, the mean and standard deviation were then calculated for every individual. Response latencies falling beyond 2.5 standard deviations from the respective individual's mean were deleted. Overall, 7.84% of the latencies were discarded.

After this data cleaning process, all subsequent analyses were conducted at the item level—that is, how fast responses were for each character across participants. To enhance reliability, we adopted a more conservative approach and analysed only items that reached at least 70% accuracy (66.67%, being 1 *SD* above chance). Out of 2500 characters, the responses for 2379 characters fulfilled this requirement and were therefore retained.

Main analysis of the seven lexical predictors using hierarchical regressions

Seven targeted lexical variables (see below for operationalization) were empirically tested for the

[2]Yang et al.'s (2009) computational model, trained on a corpus of 4468 characters, is the most extensive one thus far. However, it is a naming model. The model is also a variation of the lexical constituency model, and therefore subsequent discussion on the lexical constituency model should also apply to Yang et al. (2009).

relative importance that each plays during single-character recognition. These lexical predictors were organized into three main groups as previously described: (a) orthographic variables, (b) semantic variables, and (c) phonological variables. The definition of each predictor, together with a synopsis of previous findings, is shown in Table 1.

Orthographic variables

Frequency. Frequency represents the orthographic print exposure of a character. The subtitle (contextual diversity) frequency of Cai and Brysbaert (2010) is used here due to its superiority over other frequency measures (see Sze et al., 2014). It computes character frequency based on the number of films that a particular character appears within the subtitles (refer to Cai & Brysbaert, 2010, for a more comprehensive account). The log-values for character frequency were used in the analyses.

Number of strokes.[3] Every Chinese character comprises a unique series of stroke patterns. The number of strokes has been used to indicate a character's orthovisual complexity. The number of strokes for each character was counted and verified against the *Modern Chinese Dictionary* (The Institute of Linguistics in the Chinese Academy of Social Sciences, 2008).

Semantic variables

Number of meanings. The values used to represent polysemy were derived from counting the number of meaning entries specified for each character. The *Dictionary of Chinese Character Information* (Science Publishers, 1998; pp. 2–728) provided the basis for this metric. Although the same dictionary was consulted, these figures are not the same as the values provided by the Chinese Single-Character Word Database (i.e., the norms provided and used in the meganaming project by Y. Liu et al., 2007). This is because the figures in the Y. Liu et al. (2007) database refer to nominal categories, whereas raw counts were used in the present study.

Imageability. The imageability ratings used in our analyses were taken from the Chinese Single-Character Word Database (Y. Liu et al., 2007). A total of 480 mainland Chinese participants supplied their ratings on a 7-point Likert scale.

Age of learning (AoL). The AoL values are taken from the Chinese Single-Character Word Database, whose AoLs were compiled using the standard language textbooks published by the People's Education Press (Y. Liu et al., 2007).

Phonological variables. Although the orthography–phonology relationship appears relatively opaque for Chinese characters, phonological predictors have received considerable attention in the literature (e.g., Ziegler et al., 2000). Broadly speaking, phonological predictors can be divided into two categories: (a) those that are based on the effectiveness of the character's phonetic radical as a cue, and (b) those that are based on homophony of the character's pronunciation. To avoid multicollinearity in our analyses, consistency and phonological frequency, respectively, were selected to represent these two categories.[4] The chief difference between consistency and regularity is that regularity is necessarily categorical in nature, whereas consistency can be represented as a continuous variable. Moreover, consistency is arguably a more accurate psycholinguistic snapshot of how readers interpret characters because it might account for how nonpronounceable

[3]Number of components is another lexical variable that sought to capture the character's visuo-orthographic complexity. Components are combinations of strokes that recur across characters (see Chen & Ye, 2009, for a thorough review). However, preliminary analyses showed the variable to be highly correlated to stroke count ($r = .66$, $p < .001$). Additionally, the tolerance values for the number of strokes and number of components were very low, pitched at .536 and .554 respectively. Low tolerance is taken as statistical diagnostic of possible collinearity (Berk, 1977). To avoid multicollinearity, number of components was excluded in the final analysis. In any case, this variable produced no predictive effect when it was included as the eighth variable ($\beta = 0.205$, $p = .43$).

[4]The tolerance values for consistency and regularity were .66 and .57, respectively; and .70 and .73 for phonological frequency and homophone density, respectively. Despite the modest tolerance values for phonological frequency and homophone density, the variables are conceptually very similar; thus only one would still be selected for use (detailed explanations in the text).

phonetic components influence the pronunciation of characters. For example, although the phonetic components in characters like 杨, 炀, and 扬 are nonpronounceable (the phonetic components of these characters are located on the right), the three characters are all pronounced as /yang2/. In an important study, Fang et al. (1986) manipulated consistency within regular characters. Their results indicated that consistency was driving the regularity effects. Likewise, phonological frequency and homophone density are theoretically identical. Phonological frequency was chosen over homophone density for the present study because homophone density is not weighted by the character's frequency of occurrence, but phonological frequency is. With this greater variance, phonological frequency is likely to be the more sensitive predictor.

Consistency. Consistency reflects how systematic the mapping is between a character's phonetic component and its pronunciation. For example, 换 /huan4/ is a character with high consistency, because all other characters sharing "奂" as their phonetic component are also pronounced as /huan4/—for example, 焕, 唤, and 涣. On the other hand, 板 /ban3/ is less consistent, because characters with "反" as their phonetic component could be pronounced variously, from 皈 /gui1/, 饭 /fan4/ to 版 /ban3/, and so on. Consistency thus reflects sublexical phonological processing, given the breaking down of the character into its subparts for analysis.

At present, there are no measures available for consistency. To facilitate a fair evaluation for the effect of consistency on lexical decision latencies, a reliable metric is necessary. The computation and validation of a consistency measure is summarized in Supplemental Material 1. It should be noted that the validation is based on an independent set of naming data obtained from Y. Liu et al. (2007).[5]

Phonological frequency. Phonological frequency refers to the summed frequency of all characters sharing the same pronunciation (same onset, rime, and tone)—that is, the weighted representation of a character's homophony. This variable therefore reflects phonological information at the whole character level. The phonological frequency values in this study were taken from Chinese Single-Character Word Database (Y. Liu et al., 2007). As with print-frequency, the values for phonological frequency were log-transformed to adjust for its nonlinear distribution.

Results of the main regression analysis

Descriptive statistics. Of the possible 2379 characters that remained after data cleaning, 1560 were specified by all seven predictors and were used in the subsequent analyses. Table 2 displays the correlations among the predictors, as well as the dependent measure for these 1560 characters.

From Table 2, it is clear that most of the predictors correlated significantly and in the expected direction with mean lexical decision latency. Frequency, number of meanings, imageability, and phonological frequency were negatively related to latency, whilst stroke count and AoL were positively related. The only exception was consistency, whose effect was too small to reach significance. We revisit this point again in the discussion later. Secondly, the correlations between printed frequency and the other predictors were generally high. Strong links between frequency and other lexical properties have also been reported for Spanish (Barca, Burani, & Arduino, 2002) and English (Balota et al., 2004). Thirdly, number of strokes also correlated significantly with the rest of the predictors. These strong correlations further justify entering both frequency and stroke count as control variables in the regression analyses. Finally, AoL also shared strong relationships with the other predictors, and the pattern of correlations was in line with those reported in the literature: Characters learnt earlier tend to be of higher frequency, have fewer strokes, and are more imageable; whereas characters learnt later are usually more visuo-orthographically complex but have more consistent phonological cues (Y. Liu et al., 2007).

[5]We are grateful to Y. Liu for sharing a subset of his naming data.

Table 2. *Correlations between predictor variables and dependent variable of the 1560 characters*

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1. Mean lexical decision latency	—	−.603***	.304***	−.308***	−.100***	.462***	−.008	−.164***
2. Frequency		—	−.251***	.418***	.049†	−.555***	−.072**	.266***
3. Number of strokes			—	−.194***	.060*	.180***	.173***	−.128***
4. Number of meanings				—	−.101***	−.294***	−.017	.115***
5. Imageability					—	−.312***	−.072**	−.130***
6. Age of learning						—	.076**	−.148***
7. Consistency							—	.052*
8. Phonological frequency								—

†$p < .10$. *$p < .05$. **$p < .01$.***$p < .001$ (two-tailed).

Hierarchical multiple regression analyses of 1560 characters. Mean lexical decision latency was analysed at the item level using hierarchical regression for two reasons: This statistical procedure allowed us to test the incremental variance contributed by each group of predictors, and the predictive power of each variable can be evaluated, after the other critical predictors have been controlled for.

The predictors were entered in three stages: In Step 1, subtitle (contextual diversity) frequency and number of strokes were introduced as control variables. The reason for this is that both are typically regarded as a variables of constraint: frequency, given its robust effects and collinearity with other variables, and stroke count, as a basic index of orthovisual complexity. In Step 2, the semantic predictors—number of meanings, imageability, and AoL—were entered; and in Step 3, the phonological predictors—consistency and phonological frequency—were added.

Inserting phonological predictors last made it possible to examine how much unique variance phonology accounts for beyond the contribution of the other variables. If the universal phonological principle is correct, the phonological predictors should collectively explain a substantial amount of variance even in Step 3. In addition, at the individual level, each phonological predictor should produce a significant effect.

The results of the item-level hierarchical regression are presented in Table 3. R^2 and adjusted R^2 are both reported, and in all cases they are very similar. Standardized regression coefficients (beta, β), shown for the individual parameters, allow comparisons of the predictors in terms of their relative importance.

To investigate whether there is any substantial variation in ΔR^2 if the semantic and phonological predictors were entered differently, another set of hierarchical regression was run, with semantic predictors introduced in Step 3 and phonological predictors in Step 2. The results for collective phonological predictors ($\Delta R^2 = .006$, $p < .001$) and semantic predictors ($\Delta R^2 = .025$, $p < .001$) were similar to those in Table 3.

Discussion of results from the main hierarchical regression analysis

Orthographic variables. Orthographic variables (frequency and the number of strokes) play an important role during lexical access. When the orthographic variables were entered last in another independent hierarchical regression, they contributed an additional 16.5% variance together ($p < .001$), which is sizeable. This considerable impact of orthographic variables on character recognition was also observed when the individual orthographic predictors were examined.

Frequency. Based on the regression coefficient obtained, an increase in frequency corresponded to a significantly faster lexical decision response. This is consistent with the robust frequency effect reported for alphabetic systems and naming of Chinese characters. When the magnitudes of standardized regression coefficients were compared (Table 3), frequency was the strongest

Table 3. *Summary of results for individual predictors obtained in the final step and the incremental variance from each step of the item-level hierarchical regression*

Predictor		Regression coefficient statistics obtained at Step 3	
		β	*t*
Orthographic (constraint predictors)			
	Frequency	−0.462	-18.19^{***}
	Number of strokes	0.171	8.33^{***}
	Step 1: $\Delta R^2 = .388^{***}$ ($R^2 = .388$; $R^2_{\text{adjusted}} = .387$)		
Semantic			
	Number of meanings	−0.043	-1.98^{*}
	Imageability	−0.049	-2.31^{*}
	Age of learning (AoL)	0.155	6.17^{***}
	Step 2: $\Delta R^2 = .024^{***}$ ($R^2 = .412$; $R^2_{\text{adjusted}} = .410$)		
Phonological			
	Consistency	−0.087	-4.40^{***}
	Phonological frequency	0.007	0.32
	Step 3: $\Delta R^2 = .007^{***}$ ($R^2 = .419$; $R^2_{\text{adjusted}} = .417$)		

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

lexical predictor, after controlling for the influences from the other variables. This finding provides further evidence that frequency must be carefully accounted for in Chinese character research, and that subtitle count is a sensitive predictor.

Number of strokes. Stroke count correlated positively with lexical decision latency (Table 2). This is in line with previous work, in which shorter reaction times were recorded for characters with fewer strokes (Shen & Zhu, 1994). Number of strokes, based on a separate hierarchical regression using latencies for the 1560 characters, accounted for 2.6% ($p < .001$) more variance over and beyond the rest of the predictors, including frequency. Considering how comparatively large its beta value is when weighed against others, and the significant effect obtained (Table 3), stroke count holds substantial predictive power for character recognition by skilled readers. These results reaffirm the received view that this orthographic factor is an important consideration in a logographic script.

Semantic variables. All three semantic predictors produced reliable effects, and their results were consistent with earlier findings. Specifically, AoL was positively related to response latency, whereas the associations between latency and imageability, as well as number of meanings, were both negative. The total increment in variance by the three semantic predictors amounted to 2.4% (Table 3) beyond the joint contributions by frequency and stroke count. Furthermore, the semantic effects persisted, even after switching the order of entry with phonological variables. Taken together, these findings suggest that meaning-based information exerts an influence on the recognition of Chinese characters, although the contribution of the three specific predictors was different in nature and degree.

Number of meanings. In the present study, the effect of polysemy on decision latency was modest but significant (Table 3). This differs from the results of Y. Liu et al.'s (2007) meganaming project, where number of meanings was not a significant predictor. It could be the case that the polysemy effect in Chinese is genuinely stronger in lexical decision than in naming, but there is also the possibility that Liu et al. incurred a Type II error due to their categorization of a continuous variable. Unlike Liu et al., we preserved the continuous raw values. To test whether categorizing this continuous variable would result in a null effect, we re-ran our analysis, replacing the raw values with the nominal categories obtained from Y. Liu et al. (2007). The polysemy effect diminished and was rendered nonsignificant ($\beta = -0.033$, $p = .14$). This finding provides evidence that categorizing an inherently continuous variable, such as the number of meanings, should be avoided for a more faithful examination of its influence.

Imageability. Results from the hierarchical regression corroborate the facilitatory impact of imageability reported in extant literature (Table 3). Y. Liu et al. (2007) obtained a main effect for the naming of single characters, so our data provide converging evidence that the ease with which a referent is mentally aroused also facilitates the lexical decision for single characters.

AoL. The results in Table 3 are also consistent with existing AoL research. Larger AoL values were associated with longer lexical decision latencies ($\beta = 0.155$). More noteworthy though, is the finding that the AoL effect persisted above and beyond a strong frequency measure in Step 1. Comparing the betas in Table 3, in terms of relative importance, AoL appeared to be strongest semantic predictor, followed by imageability, then number of meanings. Thus, the age at which a character was first introduced to a child continues to have a deep impact on character processing into adulthood.

Phonological variables. When entered into the hierarchical regression at Step 3, phonological predictors explained 0.7% additional variance (Table 3). Although significant, the effect size is relatively small. This result does not seem to cohere with a strong phonological account of character recognition, such as the universal phonological principle, which would predict a robust phonological effect. Moreover, when the phonological variables were entered before the semantic variables, the outcome remained similar ($\Delta R^2 = .006$, $p < .001$). In short, the relative contribution from phonology seems less substantial than advocates of the universal phonological principle might hypothesize. Also, neither of the two phonological variables performed in the anticipated manner.

Consistency. At first glance, consistency did generate the expected pattern in our primary hierarchical analysis—that is, consistent characters are reacted to faster than less consistent characters (Table 3). However, consistency failed to yield any effect in the simple bivariate correlation (Table 2). Given that consistency did not predict lexical decision latency in isolation, but only in the presence of other lexical predictors, it seemed likely that classical suppression was at work. Classical suppression refers to the phenomenon in which a nonsignificant predictor becomes important when combined with other variables (Cohen & Cohen, 1975; Conger & Jackson, 1972; Horst, 1941; Mendershausen, 1939). Suppression effects amongst psycholinguistic variables are not well understood, but may be worthy of investigation because they appear to signify indirect lexical exchanges when characters are read.

Phonological frequency. The phonological frequency effect was not significant, after controlling for other lexical influences (Table 3). This null finding lends support to the lexical decision data reported by H.-C. Chen, Vaid, et al. (2009), who also failed to reproduce the facilitatory main effect found by Ziegler et al. (2000) using the same task. To follow up, we simulated a virtual study to examine some reasons for this lack of agreement between findings (see below). It was not possible to obtain the exact stimuli from the authors, so we decided to randomly select 28 lexical items from our pool based on the criteria set by Ziegler et al. (2000) and ran a virtual experiment to investigate what could have contributed to their finding.

Virtual experiment to clarify the phonological frequency effects

Table 4 shows that the characteristics of the stimuli in our virtual experiment and those in Ziegler et al. (2000) are generally similar. Due to the logistical difficulty of holding homophone density about 4.1 while matching for all characteristics (as Ziegler et al., 2000, reported), homophone density was set at around 14 in the virtual experiment. Notwithstanding that, homophone density was still balanced across the high and low phonological frequency conditions. Manipulation of phonological frequency is arguably more stringent in the virtual experiment, as the value for the high condition is higher than Ziegler et al.'s and lower for the low. In an effort to be consistent with Ziegler et al., statistics for printed frequency, phonological frequency, and homophone density were also computed using the same source material—that is, the *Dictionary of Modern Chinese Frequency* (Language Teaching and Research Institute of Beijing Language and Culture University, 1986). Given these control procedures, Ziegler et al.'s (2000) findings should be replicable.

Results of the virtual experiment. The results were based on latencies of correct trials from our megastudy of lexical decision. Characters with high phonological frequency were responded to faster ($M = 561.05$ ms, $SD = 59.62$) than those with low phonological frequency ($M = 579.84$ ms, $SD = 60.06$). This facilitatory main effect in phonological frequency was supported by the participant analysis, $t(34) = 2.31$, $p < .05$, $d = 0.39$, but not by item analysis, $t(26) = 1.04$, *ns*. Thus, in this virtual experiment, Ziegler et al.'s (2000) finding was, in part, replicated.

This small-scale virtual replication raises an important issue. By simply running a factorial analysis on a small sample drawn from our data, a slightly modified picture emerges. The contradiction is piquant: Information from the same dataset now provides tentative support for phonological frequency effects in lexical decision. To understand the discrepancy between results, a closer inspection of the stimuli used in the virtual experiment and a deeper examination into the nature of factorial design in conjunction with the properties of the variable are necessary.

Explaining a seemingly contradictory finding. One potential reason for the contradictory finding is the inadequate control over stimuli, an Achilles heel in some factorial manipulations. The individual characteristics of all 28 items selected for the virtual experiment are published in Supplemental Material 2. Close inspection reveals that despite the best efforts, characteristic matching can never be perfect. These accumulated imperfections could still be camouflaged strategically by comparable group means, such as those presented in Table 4. A regression analysis, on the other hand, accounts for these variations by incorporating the variables' actual values into the equation. Thus, the main effect found in the virtual study could be an artefact of these accumulated differences between conditions that persist despite the pains taken to match stimuli.

A second reason is that it is often impossible to control for all lexical variables in a factorial arrangement. Unlike the regression analysis, the virtual study failed to account for variables such as imageability and AoL. This is substantiated by H.-C. Chen, Vaid, et al.'s (2009) study, which failed to elicit any phonological frequency effect, after controlling for componential elements across the conditions.

Finally, phonological frequency is a continuous measure. In a scenario where (a) the number of items sampled is small, and (b) the original correlation between the independent and dependent

Table 4. *Comparison of stimuli characteristics in virtual experiment and Ziegler et al. (2000)*

	Condition			
	High phonological frequency		*Low phonological frequency*	
Characteristic	*Virtual expt*	*Ziegler et al.*	*Virtual expt*	*Ziegler et al.*
Mean phonological frequency (per million)	4503.5	3532	201.5	231
Mean printed frequency (per million)	102.1	101.3	102.1	101.4
Mean homophone density	14.4	4.6	14	4.1
Mean stroke count	10.6	9.6	10.1	9.7
Mean number of meanings	3.1	2.6	3.1	3.4

Note: expt = experiment.

variable is also small, dichotomizing a continuous variable will tend to augment its relationship with the dependent variable (MacCallum et al., 2002, pp. 25–26). Citing Humphreys (1978, p. 873), the "inflation of differences between means when extreme groups are used frequently inflates the experimenter's evaluation of the importance of those differences". To illustrate, we compared the effect sizes of the correlation between phonological frequency and latency, based on the 28 items in the virtual experiment, before, $r^2 = (.082)^2 = .0067$, and after dichotomization, $r^2 = (-.195)^2 = .038$. Dichotomization inflates the effect size by more than five times ($.038 \div .0067 = 5.7$), thus spuriously raising the probability of achieving a phonological frequency main effect in a parametric *t* test.

This virtual simulation highlights some of the methodological wrinkles associated with small-scale factorial studies. In the case of Chinese characters, it could explain why phonological frequency produced a significant effect in some lexical decision experiments, but not in others. Specifically, exclusive reliance on ANOVA and the attendant usage of dichotomization may skew interpretation of the processes underlying Chinese character recognition.

Examining theoretically driven interactions between variables

Beyond the individual and collective performance of lexical variables, interactions between variables can also provide crucial information about the cognitive processing of a language. In this paper, we focus on three interactions—namely, the Frequency × Number of strokes, Frequency × Polysemy, and Frequency × Consistency interactions.

The interaction between frequency and stroke count

The Frequency × Number of strokes interaction draws attention to the way orthographic complexity (stroke count) moderates lexical access. It provides some empirical grounding for a horse-race model, where holistic and componential paths contest for scarce attentional resources (see Y.-C. Chen & Ye, 2009; Taft & Zhu, 1997). In the horse-race model, when high-frequency characters are encountered, the holistic route (where the character is not parsed for analysis, and is thereby less encumbered by issues related to stroke complexity) beats the componential pathway. Low-frequency characters, by contrast, are more likely to be processed by the componential path, because they are less readily recognized and require further analytical decomposition. Logically, this decomposition process is likely to be affected by the number of strokes making up the character. For this reason, low-frequency characters with more strokes are likely to take longer to recognize than those with fewer strokes.

In essence, the horse-race model draws attention to how perceptual features, such as individual strokes, could be particularly critical in the reading of lower frequency characters. Evidence for this interaction from single-character recognition research, however, is conflicting and warrants further investigation. Shen and Zhu (1994) managed to observe a similar interaction using naming, as did Leong, Cheng, and Mulcahy (Experiment 1, 1987). Yet, neither Yu and Cao (1992, for naming) nor Peng and Wang (1997, for both naming and lexical decision) replicated the Frequency × Number of strokes interaction.

Analyses and discussion. Decision latencies to the 1560 characters were used as the dependent variables. The seven predictors used in the main hierarchical analyses were first centred. They were then entered into the regression in Step 1. The Frequency × Number of strokes interaction term was then created using the respective centred values and was then entered into the regression equation in Step 2. The interaction was found to be significant, $\beta = -0.059$, $\Delta R^2 = .003$, $p = .003$ (also see Figure 2).

Analyses of the simple slopes were conducted as follow-up. The Frequency × Number of strokes interaction observed is consistent with theoretical claims for the horse-race model, wherein a stronger stroke count effect was evident for the low-frequency characters. When frequency was set at 1 standard deviation above the mean, characters with more strokes elicited significantly slower

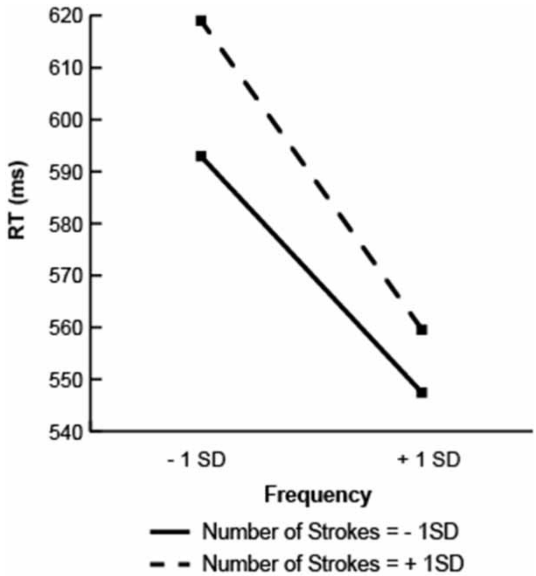

Figure 2. *Mean latency as a function of frequency and stroke count. RT = reaction time.*

responses than those with fewer strokes ($z = 3.54$, $p = .0002$). This stroke count effect is amplified when frequency was adjusted to 1 standard deviation below the mean ($z = 8.23$, $p < .00002$). Thus far, the primary evidence in support for the Frequency × Number of strokes interaction is from character naming studies (Experiment 1 in Leong et al., 1987; Shen & Zhu, 1994), so this is a novel finding for character recognition.

Interaction between frequency and the number of meanings

The second interaction explored was that of Frequency × Number of meanings. Polysemy had been found to interact with frequency using lexical decision in Chinese, at both single-character (Experiment 1 in Peng et al., 2003) and double-character word levels (Experiment 1 in B. Chen & Peng, 2001; Experiment 2 in X. Liu & Peng, 2005). Characters with more meanings elicit faster responses than those with fewer meanings, but this polysemy effect is limited to low-frequency characters. The explanation of how this interaction arises is still a matter of debate (see Borowsky & Masson, 1996; Pexman, Hino, & Lupker, 2004, for interpretations of results for English words). One possibility is that the orthographic representations of polysemous characters are activated faster, due to the additional feedback received from their multiple semantic units. However, if the connections between orthographic input layer and the orthographic output nodes required for lexical decision are stronger for high-frequency characters, there would be less opportunity for additional activations received from the semantic layer to benefit the character recognition process. In contrast, with weaker connections for low-frequency characters, semantic feedback would more influential given the longer latencies, and thus robust polysemy effects would emerge. Pertinent to the semantic account of character processing is how the Frequency × Number of meanings interaction focuses on the salience of semantic information and the readiness by which meanings can be retrieved from print in Chinese (see Lin & Ahrens, 2010, for a similar argument based on double-character Chinese words, about how polysemy effects suggest that the semantic layer is psycholinguistically salient in the mental lexicon). If semantic factors are indeed a critical influence on the lexical decision of single characters, particularly for those of lower frequency, then we should observe an interaction between the number of meanings and frequency.

Analyses and discussion. Using latencies from the same 1560 characters, the steps for the Frequency × Number of strokes were repeated. The only difference is that the Frequency × Number of meanings interaction term was created and entered in Step 2. Regression results revealed that the interaction was significant, $\beta = 0.071$, $\Delta R^2 = .004$, $p = .001$ (see Figure 3).

Further analyses on the simple slopes corroborated previous findings. When frequency was set at 1 standard deviation below the mean, characters holding more meanings were responded to significantly faster than characters with fewer meanings ($z = -3.84$, $p < .0001$). However, when frequency was set at 1 standard deviation above the mean, this polysemy effect was no longer significant ($z = 0.018$, $p = .5$). This substantial weakening of the

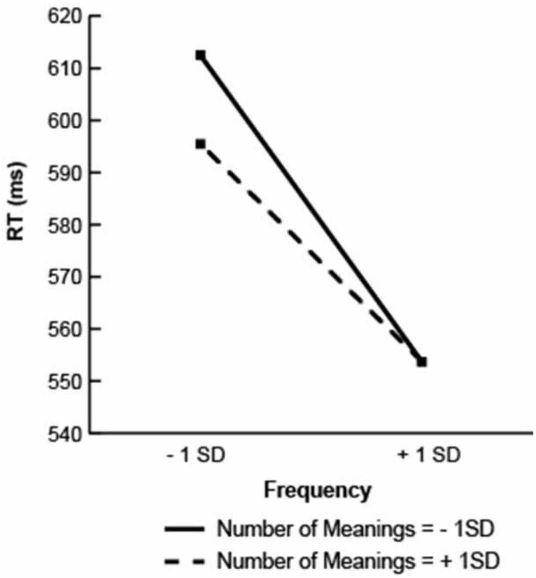

Figure 3. *Mean latency as a function of frequency and the number of meanings. RT = reaction time.*

polysemy effect across frequency reflects the typical pattern reported in literature (e.g., Peng et al., 2003). After controlling for other variables, polysemy appears to influence visual character recognition in the predicted manner, such that the influence is particularly robust if characters are low in frequency.

The interaction between frequency and consistency

The last interaction to consider is Frequency × Consistency. Seidenberg (1985) was the first to document a similar interaction between frequency and regularity in two speeded naming studies, conducted using single Chinese characters and English monosyllable words, respectively. On the basis that similar Frequency × Regularity interactions could be reproduced across Chinese characters and English words using independent samples of native-speakers, Seidenberg argued that phonological mediation is likely to occur during skilled reading of both writing systems. This idea is in keeping with an emphasis on the role of phonology when Chinese characters are read, as proposed by the phonological account. As stated, Fang et al. (1986) later clarified that it was consistency that drove the regularity effects previously found. Hue (1992) and Yang et al. (2009) then extended Seidenberg's findings by producing evidence for the Frequency × Consistency interaction using speeded naming. In this interaction, phonological consistency affects the naming of low-frequency characters, but not the naming of high-frequency items. This is due to the differences in speed, along the two hypothetical pathways, to access the pronunciation of a Chinese character. The first pathway (the direct or lexical route) allows phonology to be accessed without decomposing the character into its components. In the second pathway (the analytic route), the character is broken into its components to access its pronunciation. Specifically, the phonetic component is identified, and its neighbours (i.e., characters sharing the same phonetic component) are activated. Consistent pronunciation amongst these neighbours will generate a stronger clue on how the character is pronounced, thereby facilitating a faster response (i.e., consistency effect). For a high-frequency character, the speed of deriving a character's pronunciation is likely to be comparable between the two routes. This is because accessing the pronunciation of a frequently used character should theoretically be fast, even without having to break the character into its components (i.e., the lexical route), so having a large set of neighbours sharing the same pronunciation (thereby evoking a consistency effect along the analytic route) is not likely to facilitate a quicker response. For a low-frequency character, decomposition by way of the analytic route benefits access to its pronunciation. Presumably, the pronunciations of the phonetic component and neighbours are likely to be accessed before the whole character is lexically processed. There is, consequently, a facilitatory impact of consistency on the low-frequency characters (Hue, 1992).

Interestingly, Yang et al. (2009) developed and trained a computational model to decode Chinese characters, and it successfully simulated the Frequency × Consistency interaction. Collectively, these behavioural naming studies and the computer simulation suggest that it may be important to account for this interaction. Crucial to the phonological account is the notion that the phonetic component is prominent during the processing of a visually presented character, such that elements associated with the phonetic component (including the neighbours) are also activated. To our knowledge, the Frequency × Consistency interaction has not been examined using the lexical decision paradigm with single Chinese characters. Should phonological mediation be present even in lexical decision, we would expect consistent Chinese characters to be responded to faster than their less consistent counterparts, and that this differentiation based on consistency should be confined only to the low-frequency characters.

Analyses and discussion. Using the latencies from the same 1560 characters, the steps for the Frequency × Stroke Count interaction were repeated again, except that the Frequency × Consistency interaction term was created and entered in Step 2. Regression analysis found the interaction to be marginally significant, $\beta = -0.037$, $\Delta R^2 = .001$, $p = .057$ (see Figure 4).

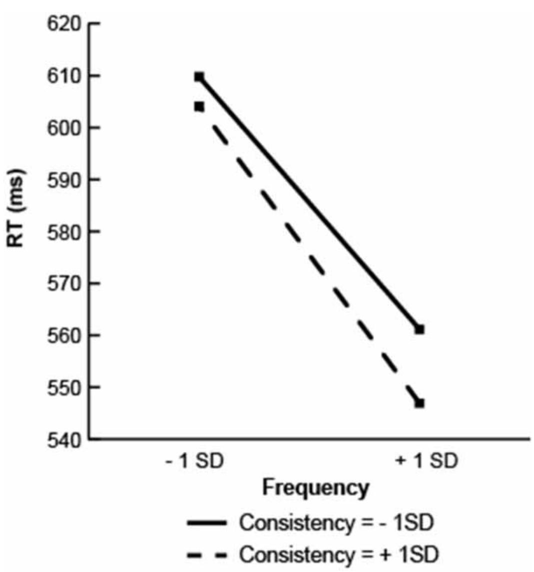

Figure 4. *Mean latency as a function of frequency and consistency. RT = reaction time.*

Follow-up analysis of simple slopes showed that when frequency was set at 1 standard deviation below the mean, consistent characters were responded to significantly faster than less consistent characters ($z = -1.76$, $p = .039$). When frequency was set at 1 standard deviation above the mean, this consistency effect was magnified ($z = -4.50$, $p < .0001$). This data pattern does not follow what is typically reported in literature (e.g., Hue, 1992), wherein stronger consistency effects should be observed for low-frequency characters. A supplementary series of analyses were also conducted. First, the characters were sorted by frequency into four equal-sized frequency quartiles. Separate multiple regressions were then run on the items within each quartile, using the same seven variables. The standardized regression coefficients obtained for consistency were −0.046 ($p = .318$; figures apply to the first quartile with low-frequency characters), −0.113 and −0.169 ($p = .027$ and $< .001$, respectively; figures apply to next two quartiles with midrange frequency characters), and, finally, −0.167 ($p < .001$; the fourth quartile with high-frequency characters). The pattern of results from this supplementary series of analyses was consistent with the analyses of simple slopes, suggesting that the underadditive Frequency × Consistency interaction observed was not likely to be an artefact induced by a small set of idiosyncratic items. A straightforward interpretation of this interaction should be cautioned against for two reasons. First, the main effect of consistency itself is qualified by the presence of classical suppression. Secondly, the interaction is not very robust, and in fact only marginally significant ($p = .057$). Therefore, it would be prudent to follow up on this finding in future megastudies that employ other character-processing tasks (including speeded naming). This will help verify whether the underadditive interaction obtained here is specific to lexical decision. It is also possible that this underadditive interaction is a consequence of character construction. Low-frequency characters tend to be more consistent, in that they are more likely to share the same pronunciation as their phonetic stems. This is supported by the negative relationship observed between frequency and consistency ($r = -.06$, $p = .03$). If so, there is an argument to be made that pronunciation of a low-frequency character is less likely to be affected by the phonology of its neighbours, since readers will simply pronounce it by way of its phonetic component, ignoring other plausible pronunciations. The reverse applies to high-frequency characters. For this reason, consistency exerts a greater influence on high-frequency characters (the less consistent characters), prompting readers to consider alternative pronunciations based on their constituent phonetic components. There is also a possibility that high-frequency characters may have enemies (characters sharing the same phonetic stem but pronounced differently) whose character frequencies are similarly high. This in turn drives consideration of alternative pronunciations.[6] Nevertheless, these are post hoc and speculative explanations, and future research on character construction and the corpus could shed light on their plausibility. In general, although there is good evidence for phonological consistency (and Frequency × Consistency interaction) in naming, its robustness is qualified by naming's emphasis on articulatory processes (and thus computation of phonology). In a task where the computation of phonology is not obligatory, evidence for consistency and its associated interaction(s) might not to be as clear-cut. Lastly, it should be reemphasized that to our knowledge, no study to date has examined the Frequency × Consistency interaction in lexical decision for Chinese single-character recognition. Whether the present pattern is robust or whether it is driven by the task-specific demands of lexical decision are important questions that merit future investigation.[7] Interestingly, the joint effects of consistency and frequency seem to manifest themselves differently in speeded naming and lexical decision, even in English. Although there is evidence that consistency and frequency interact overadditively in English factorial *naming* studies, the same interaction does not hold up in factorial lexical decision studies and in megastudy data (see Keuleers et al., 2012). In short, there is evidence, even in English, that the overadditive pattern in lexical decision is not observed.

GENERAL DISCUSSION

This paper describes a megastudy analysis of orthographic, semantic, and phonological variables on the visual recognition of Chinese single characters using lexical decision latencies. Specifically, attention was paid to the roles of seven established predictors (frequency, number of strokes, number of meanings, imageability, AoL, consistency, and phonological frequency). The use of a large pool of stimuli (1560 single characters) and hierarchical regression provided a means of avoiding the problems that typically arise from small-scale factorial studies, which include the inadequate control of confounding variables and the possibility of overemphasizing subtler effects through selected stimuli (Spieler & Balota, 1997).

Thus far, while there is general acceptance for the role performed by the orthographic variables, considerable debate ensues regarding semantic and phonological variables. The traditional view is that visual recognition of characters was purely mediated by semantics (Hoosain, 1991; W. S.-Y.

[6]We thank an anonymous reviewer for suggesting the second possibility.

[7]An additional 2 × 2 within ANOVA analysis was run, based on the same factorial set-up and stimuli as those used in Study 2 of Yang et al. (2009). Character frequency (middle/low) was crossed against consistency (consistent/inconsistent) for 44 characters that were common between Yang et al.'s naming study and the Chinese Lexicon Project. The results did not reveal a Frequency × Consistency interaction [$F(1, 34) = 2.25$, $p = .14$, $MSE = 2923.41$, $\eta^2_{\text{partial}} = .062$]. However, the descriptive statistics from this small-scale factorial reanalysis yielded a similar trend to the underadditive pattern we have observed, such that the reaction times taken for the low-frequency items were similar (low-frequency, consistent characters: $M = 682.74$, $SD = 91.43$; low-frequency, inconsistent characters: $M = 678.48$, $SD = 99.32$), but a greater discrepancy in speed was noted for the middle-frequency items. Reaction for the middle-frequency, consistent characters ($M = 609.65$, $SD = 71.19$) appeared slightly faster than that for the middle-frequency, inconsistent characters ($M = 632.79$, $SD = 79.40$). There is therefore some possibility that the underadditive interaction obtained in this paper might even be potentially reproducible in factorial studies using lexical decision. The authors would like to thank J. Yang and Jason Zevin for sharing their stimuli.

Wang, 1973), but this was challenged subsequently by the phonological account (Perfetti & Dunlap, 2008; Tan & Perfetti, 1997). A strong form of the phonological account was the universal phonological principle, which stated that all languages, including Chinese, are phonologically mediated (Frost, 1998; Ziegler et al., 2000). Here, the balance between orthography, semantics, and phonology was reexamined from the vantage point of a megastudy.

Lexical phenomena that single-character recognition model(s) should be able to address

The role of orthographic variables

In the main analysis using hierarchical regression, the unique predictive power of each lexical variable was compared. The two orthographic variables, frequency and stroke count, together explained considerable variance (16.5% additional variance when they were entered at Step 3, after the semantic and phonological variables). At the individual level, frequency ($\beta = -0.462$) and number of strokes ($\beta = 0.171$), were the two strongest predictors, after controlling for the other variables. These findings corroborate the received view that frequency and stroke information are both fundamental to character recognition by skilled readers and reinforce the need to account for these two orthographic predictors in future work.

The reported interaction between frequency and number of strokes (Shen & Zhu, 1994) was also replicated, such that low-frequency characters display stronger stroke effect than their high-frequency counterparts. This result clarifies the discrepancies between findings previously obtained by small-scale factorial studies (e.g., the conflicting findings between Shen & Zhu, 1994, and Peng & Wang, 1997). Future computational instantiations should be able to accommodate this horse-race phenomenon related to how frequency seems to govern the salience of orthographic (stroke) analysis.

The role of semantic variables

Semantic variables (number of meanings, imageability, AoL) also accounted for a significant amount of variance, both collectively and individually during the lexical decision task. Collective performance remained strong, even when semantic factors were entered last (2.5% additional variance) after the orthographic and phonological variables. Additionally, the number of meanings also appeared to significantly influence the lexical decision of low-frequency characters, as revealed by the Frequency × Number of meanings interaction. These findings are in tandem with the view that for a logographic script, meaning-based information critically affects visual recognition, and endorses the view that character recognition involves semantic information.

There are strong theoretical reasons to expect semantics to play an important role. First, phonetic stems only reliably cue the character's pronunciation in just 29% of all possible semantic–phonetic compound characters (Y. Zhou, 2003, p. 85). Secondly, script-to-sound mapping in Chinese tends to be more arbitrary than script-to-meaning correspondence. Though not always obvious, most characters including the phonograms do have semantic origins (Gu, 2007). Thirdly, task demands could have potentially skewed previous findings. The general dependence on naming as the experimental task for phonological evidence is one such example. Williams and Bever (2010) tried to find evidence on how an experiment's paradigm could induce a particular processing strategy by conducting three separate experiments. The first was a semantic categorization task, and they found that compound characters with semantically relevant radicals were grouped faster. The second was a homonym recognition task, in which characters with the same pronunciation and same phonetic stems were recognized faster than characters with different pronunciations but different phonetic stems. These two experiments served to illustrate that assigning participants to complete either a semantic or phonological task will draw the corresponding semantic or phonological effects. In the final experiment, Williams and Bever employed a strategy-neutral task. They designed a lexical decision task, such that at any one time, either the semantic or phonetic component of the displayed target was masked with a Gaussian blur.

Their rationale was that if Chinese readers were inherently semantically biased, when semantic information is degraded, their responses will be slowed down, and vice versa for phonology. They found a significant impairment effect for conditions with degraded semantic components, suggesting that in a neutral situation, Chinese readers were more inclined to rely on semantic information. In short, Williams and Bever provided evidence against the phonological account. Their results also emphasized the important role of the experimental task in directing specific strategies.

In this paper, the lexical decision task implemented does not involve the use of explicit or implicit (e.g., priming) semantic or phonological prompts. The variance accounted for by the semantic variables in this paper is thus more consistent with observations that in the absence of deliberate cueing, readers tend to interpret or analyse Chinese characters semantically (Y. Liu et al., 2007; Williams & Bever, 2010).

In terms of theoretical models, there is still no specific computational model with features that highlight the premium on the roles performed by semantic variables during character reading. Future computational models should be able to account for the sizeable variance predicted by semantic predictors in single-character recognition, as well as the interaction between polysemy and frequency.

The role of phonological variables

The universal phonological principle posits that the core representations within the mental lexicon are acoustic codes. The corollary is that strong contributions from phonological predictors would be expected, even in lexical decision, which does not require explicit phonological recoding (but see Halderman et al., 2012, on the usefulness of lexical decision as a paradigm to reveal early phonology). However, our results from the main hierarchical regressions failed to support the strong form of the phonological account. Collectively, the phonological variables explained only 0.7% additional variance when entered last, compared to the 2.5% additional variance for semantic variables or 16.5% for orthographic variables. Importantly, the magnitude of the contribution of phonology still remained relatively minor, even when the phonological predictors were entered into the equation before the semantic variables. At the individual level, phonological frequency failed to produce any observable effect, after controlling for other variables. Consistency effects were significant, but the effect size is small ($\beta = -0.087$). The analyses also did not reveal a clear-cut facilitatory effect for consistency because classical suppression was observed. Further work is necessary to clarify the role of consistency in character recognition. Phonological frequency and consistency reflect phonological information accessed at the whole-character and sublexical (decomposition of character into its stems) level, respectively, thus considering their results in combination should be representative of character processing. Overall, the results achieved are generally in line with Y. Liu et al.'s (2007) megastudy of naming data, in which they stated in unequivocal terms that "the effect of phonology on [Chinese] word naming is not significant" (p. 196). Pending clarification on the results of consistency, a more moderate stand on the role of phonology seems appropriate. However, the failure to observe statistically powerful phonological effects in both megastudies carried out independently by different laboratories does mitigate against a strong claim for the phonological account.

Evaluating the multilevel interactive activation framework and lexical constituency model

There are two theoretical models that target the reading of Chinese characters—namely, the multilevel interactive activation framework (Taft & Zhu, 1997; Taft, Liu, & Zhu, 1999) and the lexical constituency model (Perfetti & Liu, 2006). Both models follow McClelland and Rumelhart's (1981) connectionist tradition. Consequently, in both these models, the orthography–phonology–semantics triangle is embedded within the architectures as core layers. The primary differences between the two models rest on the details related to how each layer might be further segmented and connected.

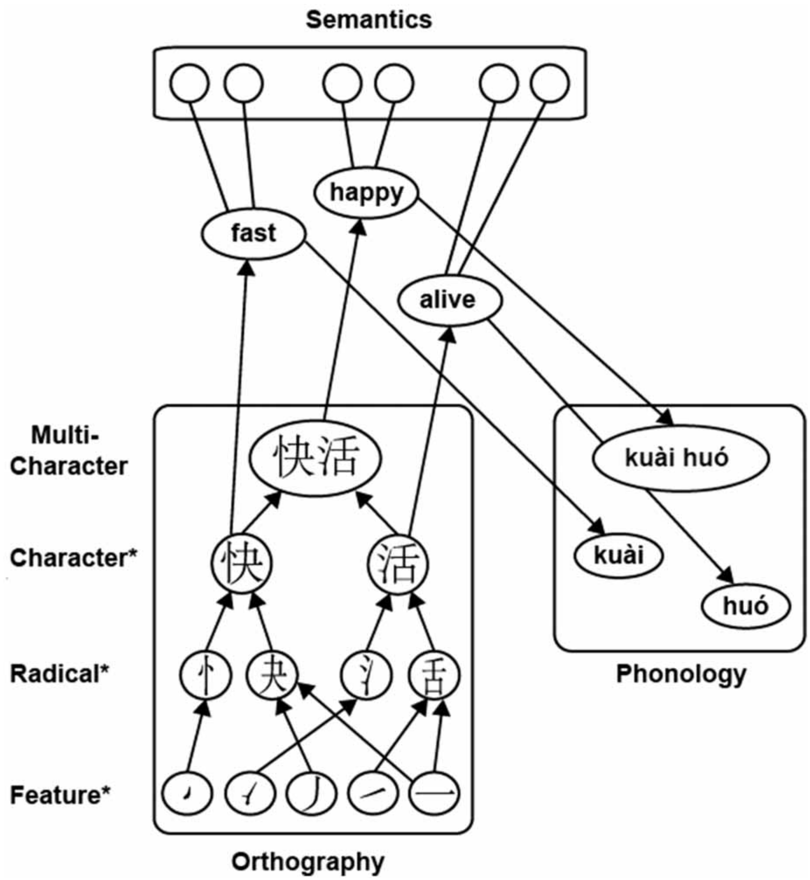

Figure 5. *The multilevel interactive activation framework.*

The multilevel interactive activation framework

The framework was initially designed to account for small-scale lexical decision data (Taft & Zhu, 1997) and is therefore most pertinent to our discussion. The full model seeks to include information from the stroke level onwards to the word (double-character) level (Taft, 2006; Taft, Liu, & Zhu, 1999). For our purposes, the components and connections below the multicharacter level will be relevant (marked with asterisk in Figure 5). The main contribution of this model is the unpacking of the orthographic layer into stroke, radical, and character (morpheme) components. Megastudy evidence corroborating the presence of stroke and morpheme components is supplied by the respective main effects of stroke and frequency found in this paper. However, a key feature of their model is the proposition that radicals in Chinese characters are position-specific, and the researchers provided evidence based on the position-specific radical frequency (Taft, Zhu, & Peng, 1999). Unfortunately, we were unable to include this variable within our analyses due to a lack of a reliable metric.[8] One possible source was a linguistic project jointly supported by the Ministry of Education of the People's Republic of

[8]The metric used by the researchers in support of the multilevel interactive activation framework is published in a *Chinese Radical Position Frequency Dictionary* (1984). However, there might be only one hand-written copy in existence, and this will not be accessible to most researchers.

China and the State Language Commission (2009), but the metric has yet to be properly compiled for release (H. Xing, personal communication, July 10 2012).[9] Should a resource be made available in the future, it will be interesting to examine whether a main effect of the variable would be observed, after controlling for other predictors.

As the multilevel interactive activation framework is rather general and lacks explicit details about the connections between components or layers, it can only be assumed at this stage that the various lexical phenomena, such as the Frequency × Number of strokes and Frequency × Number of meanings interactions could be accommodated by the framework through potential adjustments of weights between nodes. However, the litmus test of its viability would only be when the framework is computationally implemented, thus enabling, beyond theoretical discussion, concrete evaluation with behavioural data, such as ours.

The lexical constituency model

This model was designed to explain primed character naming in relation to time course activation. However, aside from differences in output (the output component would presumably deals with phonological assembly for naming purposes), its basic architecture for decoding characters could be extended to other paradigms, including lexical decision. The architecture (see Figure 6) follows the typical orthography–phonology–semantics layout, which was previously described. The authors have posited specific predictions about how the lexical constituency model could account for particular lexical effects (Perfetti & Tan, 1999). Three are relevant for our discussion.

First, the constituency principle, the foundation of the model, states that all three constituents (orthography, phonology, and semantic) will be activated to different extents when characters are read. This is in line with our data, as all three types of variables yielded significant effects.

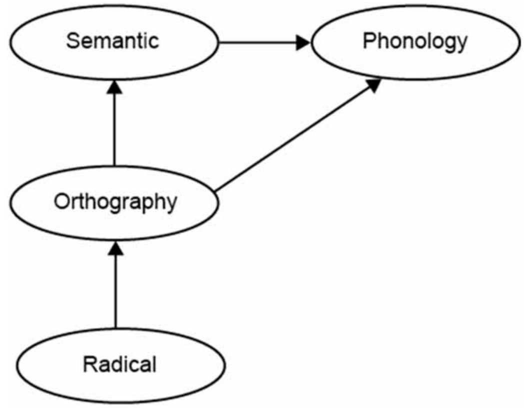

Figure 6. *The lexical constituency model.*

Secondly, Perfetti and Tan (1999) emphasized the role of phonological mediation during character identification. Specifically, they argued that "frequency-driven phonological effect is regarded as another basis for early phonological activation" (Perfetti & Tan, 1999, p. 129; Tan & Perfetti, 1998). They also developed a theoretical account of how phonological units are linked to meaning units: For characters with fewer homophones, phonological nodes should influence access to character meanings to a greater extent than for characters with many homophones (Tan & Perfetti, 1997). Importantly, this emphasis on phonological mediation during access to meaning based on phonological frequency is not supported by our megastudy data. We failed to detect any effects for phonological frequency during lexical decision, and yet there were semantic effects on latencies. In general, we found that phonological effects are relatively weak and are thus unlikely to wield the considerable influence on character recognition that Perfetti and Tan's model would predict.

Thirdly, Perfetti and Tan (1999, p. 130) asserted that semantic uncertainty will influence character recognition, as this vagueness affects how units diverge on semantic units. There is support for this theoretical claim, as a main effect for the number of meanings was present in our data.

[9]Xing was one of the primary researchers involved in the project.

Similar to the multilevel interactive activation framework, the connections in the lexical constituency model are not provided in detail. This is acknowledged in their statement "we judge the evidence insufficient to postulate such connections in detail, but they can be added as evidence requires" (Perfetti & Tan, 1999, p. 131). This makes a full and fair assessment of the model difficult, but in theory, it could accommodate most of the lexical phenomena benchmarked by our data, through future weight adjustment. Interestingly, there has been a rudimentary attempt to implement the lexical constituency model computationally (Perfetti & Liu, 2006; Perfetti et al., 2005), but it is not yet possible to cross-check the model's functionality with our behavioural data. The reasons are that (a) its computerized version recognizes a very small corpus of characters (204 characters), (b) its connection weights were fixed, not trained, so this hardwired model cannot recognize new characters beyond the 204 characters, and (c) the targeted paradigm was primed character naming. Analogous models for English capable of effective evaluation, such as the parallel distributed processing (PDP) model and connectionist dual process (CDP+) model, were trained on 2897 monosyllabic words (Seidenberg & McClelland, 1989) and 7383 unique orthographic patterns (Perry et al., 2007), respectively, allow free weight adjustments to simulate learning, and could explain at least both word and nonword naming (the PDP model explains lexical decision data as well). Nonetheless, it is probably only a matter of time before a computerized form of a more fully developed and versatile Chinese character recognition model will be realized for research evaluation.

The utility of the megastudy approach

Throughout various phases in our analyses, the usage and advantages of the megastudy approach were demonstrated. It allows researchers to validate new normative measures, as in the case of a consistency metric (Supplemental Material 1), and the data can be resampled to run virtual experiments. In the follow-up analyses, a virtual experiment was conducted to clarify why findings related to phonological frequency were contradictory. Some outstanding experimental issues were brought to attention. These included the logistical difficulty of matching stimuli across experimental cells in a factorial study, as well as the cost of dichotomizing continuous variables. Many findings on Chinese characters were previously corroborated using replications, but if these replications were also based on factorial designs, similar distortions and potential bias in the stimuli could be operating. Additionally, the effects related to item selection in factorial studies do raise serious concern, as results based on a restricted number of carefully sampled stimuli might not generalize to other characters. The discrepancies highlighted in this paper make it clear that factorial studies should be complemented with research based on other statistical approaches, such as the megastudy. A more extensive set of stimuli is more representative of the reader's lexicon. The continued clarification and corroboration of findings obtained from small-scale factorial designs is imperative for theoretical advancement.

Future directions

At least three directions can be identified for future research. First, seven predictors with available and verifiable measures were evaluated in this paper. This pool of selected predictors is not exhaustive: Component frequency (or radical frequency), for instance, could be explored, but there are no publicly accessible and reliable norms released for this variable. Nonetheless, performance of this variable, particularly its position-specific variant (i.e., position-specific radical frequency count), seems to be the lynchpin of the multilevel interactive model (Taft & Zhu, 1997), so examining some of the model's predictions seems warranted, were a suitable measure to be made available.

Secondly, in the present lexical decision megastudy, as well as in Y. Liu et al.'s (2007) meganaming project, the stimuli used were single characters. Single characters are the most basic and holistic lexical unit that can concatenate Chinese words. Thus, findings from single characters will be fundamental and instrumental. In saying so, the Chinese lexicon is primarily composed of compound words—that is, multiple-character words (double characters or more). 现代 /xian4 dai4/, for example, is a double-character word that means "modern", and it is made up of two characters, "现" and "代". It will be tremendously useful to extend the megastudy approach to explore the lexical processes involved when these multiple-character words are read.

Finally, the characteristics of the participants making up the megastudy are likely to have considerable influence on the data generated. In this megastudy, lexical decision data were collected from mainland Chinese participants, who are native-speakers of Mandarin and skilled readers of the simplified form of Chinese characters. The reason for starting with this participant pool is because these first-language users provide a clear basis to understand how Chinese characters are cognitively processed. There are, however, many types of Chinese language users, with other language profiles, whose first language may not necessarily be Chinese or they may speak another dialectal form of Chinese, aside from Mandarin. In Hong Kong, for instance, the Chinese participants are likely to speak Cantonese and use the traditional complex characters. It remains unclear how these script and language differences could affect lexical processes in character reading. Cantonese, for instance, has nine tones, compared to Mandarin's four basic tones. Will the Hong Kong Cantonese sample be more sensitive to phonology? A megastudy based on Cantonese speakers in Hong Kong will be an effective way to address these questions. In Singapore, a major segment of the Chinese population is English educated and speaks Mandarin only as the second language. Will these subordinate bilinguals process Chinese characters differently? A similar megastudy based on Singaporean English–Chinese speakers and other Chinese-knowing bilinguals would be informative. By comparing the results of megastudies for these different Chinese language samples against a native control group, such as ours, a more complete picture of Chinese character processing will be yielded. Only then can researchers make more definitive statements related to common or "universal" principles specific to Chinese character processing.

To conclude, our study leverages on the megastudy approach to examine the visual recognition of single Chinese characters. It provides converging evidence for several main effects, notably the influences of orthographic and semantic variables. These include the effects of frequency, stroke count, number of meanings, imageability, and AoL. Although not all predictors could be encompassed within the scope of the current study, the critical variables and whose veracity could be attested for were included in the regression analyses. It is a modest start, and hopefully an invaluable start. A balanced understanding of visual word recognition, based on a larger and more representative character sample, should be of benefit to the psycholinguistic research enterprise.

Supplemental material

Supplemental content is available via the "Supplemental" tab on the article's online page (http://dx.doi.org/10.1080/17470218.2014.985234).

Original manuscript received 21 January 2014
Accepted revision received 18 October 2014

REFERENCES

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. doi:10.1037/0096-3445.133.2.283

Balota, D. A., & Spieler, D. H. (1998). The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science*, *9*(3), 238–240. doi:10.1111/1467-9280.00047

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: Large scale analysis of lexical processes. In J. S. Adelman (Ed.), *Visual word recognition volume 1: Models and methods,* *orthography and phonology* (pp. 90–115). Hove, East Sussex: Psychological Press.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 424–434. doi:10.3758/BF03195471

Berk, K. N. (1977). Tolerance and condition in regression computations. *Journal of the American Statistical Association*, *72*(360), 863–866.

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 63–85. doi:10.1037/0278-7393.22.1.63

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE 5*(6), e10729. doi:10.1371/journal.pone.0010729

Chen, B., Dent, K., You, W., & Wu, G. (2009). Age of acquisition affects early orthographic processing during Chinese character recognition. *Acta Psychologica*, *130*(3), 196–203. doi:10.1016/j.actpsy.2008.12.004

Chen, B., & Peng, D. (2001). 汉语双字多义词的识别优势效应 [The effects of polysemy in two-character word identification]. *Acta Psychologica Sinica*, *33*(4), 300–304.

Chen, B., You, W., & Zhou, H. (2007). 汉语词汇习得的年龄效应: 语义假设的证据 [Age of acquisition effects in reading Chinese : Evidence in favor of the semantic hypothesis]. *Acta Psychologica Sinica*, *39*(1), 9–17.

Chen, B. G., Zhou, H. X., Dunlap, S., & Perfetti, C. A. (2007). Age of acquisition effects in reading Chinese: Evidence in favour of the arbitrary mapping hypothesis. *British Journal of Psychology*, *98*(3), 499–516. doi:10.1348/000712606X165484

Chen, H.-C., & Shu, H. (2001). Lexical activation during the recognition of Chinese characters: Evidence against early phonological activation. *Psychonomic Bulletin & Review*, *8*(3), 511–518. doi:10.3758/BF03196186

Chen, H.-C., Vaid, J., & Wu, J.-T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language and Cognitive Processes*, *24*(7-8), 967–982.

Chen, Y.-C., & Ye, S.-L. (2009). 汉字辨识理论模型中的部件表徵 [Radical representation in models of Chinese character recognition]. *Research in Applied Psychology*, *43*, 177–205. doi:10.1080/01690960902804515

Chiang, W.-C. (2003). The basic orthographic units in children's recognition of Chinese characters. *Chinese Journal of Psychology*, *45*(1), 47–71.

Chung, C. M., Cheng, S. C., & Leong, C. K. (1988). Interaction of reader and stimulus characteristics in children's recognition of Chinese. In I.-M. Liu, H.-C. Chen, & M. J. Chen (Eds.), *Cognitive aspects of the Chinese language, Vol. 1* (pp. 41–56). Hong Kong: Asian Research Service.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253. doi:10.1177/014662168300700301

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). London: Academic Press.

Coltheart, M. (2012). Dual-route theories of reading aloud. In J. S. Adelman (Ed.), *Visual word recognition volume 1: Models and methods, orthography and phonology* (pp. 3–27). Hove, East Sussex: Psychological Press.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. doi:10.1037/0033-295X.108.1.204

Conger, A. J., & Jackson, D. N. (1972). Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and Psychological Measurement*, *32*(3), 579–599. doi:10.1177/001316447203200303

Fang, S. P., Horng, R.-Y., & Tzeng, O. J. L. (1986). Consistency effects in the Chinese character and pseudo-character naming tasks. In H. S. R. Kao & R. Hoosain (Eds.), *Linguistics, phonology, and the Chinese language* (pp. 11–21). Hong Kong: Centre of Asian Studies, University of Hong Kong.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*(1), 71–99. doi:10.1037//0033-2909.123.1.71

Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica*, *115*(1), 43–67. doi:10.1016/j.actpsy.2003.11.002

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research*

Methods & Instrumentation, *12*(4), 395–427. doi:10.3758/BF03201693
Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 674–691. doi:10.1037/0096-1523.5.4.674
Gu, J.-P. (2007). 字解: 字形图解字典 [A compendium of Chinese characters]. Singapore: Chinese Heritage Lodge.
Halderman, L. K., Ashby, J., & Perfetti, C. A. (2012). Phonology: An early and integral role in identifying words. In J. S. Adelman (Ed.), *Visual word recognition volume 1: Models and methods, orthography and phonology* (pp. 207–228). Hove, East Sussex: Psychological Press.
Han, Z., Bi, Y., Shu, H., & Weekes, B. S. (2005). The interaction between semantic and sublexical routes in reading: Converging evidence from Chinese. *Brain and Language*, *95*(1), 235–236. doi:10.1016/j.bandl.2005.07.120
Honorof, D. N., & Feldman, L. (2006). The Chinese character in psycholinguistic research: Form, structure, and the reader. In P. Li (Series and Vol. Ed.), L. H. Tan, E. Bates, & O. J. L. Tzeng (Vol. Eds.), *Handbook of East Asian psycholinguistics: Vol. 1. Chinese* (pp. 237–249). Cambridge, UK: Cambridge University Press.
Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Horst, P. (1941). *The prediction of personal adjustment*. New York: Social Science Research Council.
Hue, C. W. (1992). Recognition processes in character naming. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 93–107). Amsterdam: Elsevier North-Holland.
Humphreys, L. G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. *Journal of Educational Psychology*, *70*(6), 873–876. doi:10.1037/0022-0663.70.6.873
Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304. doi:10.3758/s13428-011-0118-4
Language Teaching and Research Institute of Beijing Language and Culture University. (1986). 现代汉语频率词典 [Dictionary of modern Chinese frequency]. Beijing: Beijing Language and Culture University Press.
Leong, C. K., Cheng, P.-W., & Malcahy, R. (1987). Automatic processing of morphemic orthography by mature readers. *Language and Speech*, *30*(2), 181–196. doi:10.1177/002383098703000207
Li, C. N., & Thompson, S. A. (2009). Chinese. In B. Comrie (Ed.), *The world's major languages* (2nd ed., pp. 703–723). Abingdon: Routledge.
Lichacz, F. M., Herdman, C. M., Lefevre, J.-A., & Baird, B. (1999). Polysemy effects in word naming. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *53*(2), 189–193. doi:10.1037/h0087309
Lin, C.-J. C., & Ahrens, K. (2010). Ambiguity advantage revisited: Two meanings are better than one when accessing Chinese nouns. *Journal of Psycholinguistic Research*, *39*(1), 1–19. doi:10.1007/s10936-009-9120-8
Liu, I.-M. (2003). Introduction to Chinese character/word processing. *Chinese Journal of Psychology*, *45*(1), 1–9.
Liu, I.-M., Wu, J.-T., & Chou, T.-L. (1996). Encoding operation and transcoding as the major loci of the frequency effect. *Cognition*, *59*(2), 149–168. doi:10.1016/0010-0277(95)00688-5
Liu, X., & Peng, D. (2005). 词汇判断中汉语多义词识别的优势效应 [The effects of polysemy for Chinese words in lexical decision]. *Studies of Psychology & Behavior*, *3*(2), 116–120.
Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, *39*(2), 192–198. doi:10.3758/BF03193147
MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40. doi:10.1037/1082-989X.7.1.19
McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. doi:10.1037/0033-295X.88.5.375
Mendershausen, H. (1939). Clearing variates in confluence analysis. *Journal of the American Statistical Association*, *34*(205), 93–105. doi:10.1080/01621459.1939.10502370
Ministry of Education of the People's Republic of China & State Language Commission. (2009). 现代常用字部件及部件名称规范 [Specification of common modern Chinese character components and component names]. Retrieved June 30, 2012, from http://www.china-language.gov.cn/9/2009_9_27/1_9_4352_0_1254018842817.html

Peng, D., Deng, Y., & Chen, B. (2003). 汉语多义单字词的识别优势效应 [The polysemy effect in Chinese one-character word identification]. *Acta Psychologica Sinica*, *35*(5), 569–575.

Peng, D. & Wang, C. (1997). *汉字加工的基本单元*: 来自笔画数效应和部件数效应的证据[Basic processing unit of Chinese character recognition: Evidence from stroke number effect and radical number effect]. *Acta Psychologica Sinica*, *29*(1), 8–16.

Perfetti, C. A., & Dunlap, S. (2008). Learning to read: General principles and writing system variations. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second-language literacy development* (pp. 13–38). Mahwah, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A., & Liu, Y. (2006). Reading Chinese characters: Orthography, phonology, meaning, and the Lexical Constituency Model. In P. Li (Series and Vol. Ed.), L. H. Tan, E. Bates, & O. J. L. Tzeng (Vol. Eds.), *Handbook of East Asian psycholinguistics: Vol. 1. Chinese* (pp. 225–236). Cambridge, UK: Cambridge University Press.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The Lexical Constituency Model: Some implications of research on Chinese for general theories of reading. *Psychological Review*, *112*(1), 43–59. doi:10.1037/0033-295X.112.1.43

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 101–118. doi:10.1037/0278-7393.24.1.101

Perfetti, C. A., & Tan, L. H. (1999). The Constituency Model of Chinese word identification. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 115–134). Mahwah, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a 'universal' phonological principle. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227–248). Amsterdam: Elsevier North-Holland.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315. doi:10.1037/0033-295X.114.2.273

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1252–1270. doi:10.1037/0278-7393.30.6.1252

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115. doi:10.1037/0033-295X.103.1.56

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 487–494. doi:10.1016/S0022-5371(70)80091-3

Science Publishers. (1998). 汉字信息字典 [A Dictionary of Chinese character information]. Shanghai: Author.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*(1), 1–30. doi:10.1016/0010-0277(85)90029-0

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. doi:10.1037/0033-295X.96.4.523

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word-reading models at the item level: Matching the grain of theory and data. *Psychological Science*, *9*(3), 234–237. doi:10.1111/1467-9280.00046

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, *134*(4), 584–615. doi:10.1037/0033-2909.134.4.584

Shen, L. M., & Zhu, X. (1994). 汉字识别中笔画数与字频效应的研究 [Investigating stroke count and character frequency effects in the visual recognition of characters]. *Psychological Science*, *17*(4), 245–247.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*(6), 411–416. doi:10.1111/j.1467-9280.1997.tb00453.x

Sun, C. (2006). *Chinese: A linguistic introduction*. New York: Cambridge University Press.

Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods*, *46*, 263–273. doi:10.3758/s13428-013-0355-9

Taft, M. (2006). Processing of characters by native Chinese readers. In P. Li (Series and Vol. Ed.), L. H. Tan, E. Bates, & O. J. L. Tzeng (Vol. Eds.), *Handbook of East Asian psycholinguistics:*

Vol. 1. Chinese (pp. 237–249). Cambridge, UK: Cambridge University Press.

Taft, M., Liu, Y., & Zhu, X. (1999). Morphemic processing in reading Chinese. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 91–113). Mahwah, NJ: Lawrence Erlbaum Associates.

Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experiment Psychology: Language, Memory, and Cognition*, *23*(3), 761–775. doi:10.1037/0278-7393.23.3.761

Taft, M., Zhu, X., & Peng, D. (1999). Positional specificity of radicals in Chinese character recognition. *Journal of Memory and Language*, *40*(4), 498–519. doi:10.1006/jmla.1998.2625

Tan, L. H., & Perfetti, C. A. (1997). Visual Chinese character recognition: Does phonological information mediate access to meaning? *Journal of Memory and Language*, *37*(1), 41–57. doi:10.1006/jmla.1997.2508

Tan, L.-H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. *Reading and Writing: An Interdisciplinary Journal*, *10*(3–5), 165–200. doi:10.1023/A:1008086231343

The Institute of Linguistics in the Chinese Academy of Social Sciences. (2008). 现代汉语词典 [Modern Chinese dictionary] (5th ed.). Beijing: The Commercial Press.

Wang, K. (2011). An electrophysiological investigation of the role of orthography in accessing meaning of Chinese single-character words. *Neuroscience Letters*, *487*(3), 297–301. doi:10.1016/j.neulet.2010.10.041

Wang, K., Mecklinger, A., Hofmann, J., & Weng, X. (2010). From orthography to meaning: An electrophysiological investigation of the role of phonology in accessing meaning of Chinese single-character words. *Neuroscience*, *165*(1), 101–106. doi:10.1016/j.neuroscience.2009.09.070

Wang, W. S.-Y. (1973). The Chinese language. *Scientific American*, *228*(2), 50–60.

Williams, C., & Bever, T. (2010). Chinese character decoding: A semantic bias? *Reading and Writing: An Interdisciplinary Journal*, *23*(5), 589–605. doi:10.1007/s11145-010-9228-0

Yang, J., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language*, *61*(2), 238–257. doi:10.1016/j.jml.2009.05.001

Yin, B., & Rohsenow, J. S. (1994). *Modern Chinese characters*. Beijing: Sinolingua.

Yu, B., & Cao, H. (1992). 汉字识别中的笔画数效应新探兼论字频效应 [A new exploration on the effect of stroke number in the identification of Chinese characters]. *Acta Psychologica Sinica*, *24*(2), 120–126.

Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1245–1265. doi:10.1037/0278-7393.26.5.1245

Zhou, Y. (2003). *The historical evolution of Chinese languages and scripts* (L. Zhang, Trans.). Columbus, OH: Ohio State University National East Asian Languages Resource Center. (Original work published in 1997).

Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology matters: The phonological frequency effect in written Chinese. *Psychological Science*, *11*(3), 234–238. doi:10.1111/1467-9280.00247