CHARACTER RECOGNITION WITHOUT SOUND OR MEANING[1]

James Myers*, Marcus Taft, and Peiying Chou***

**National Chung Cheng University and **University of New South Wales*

ABSTRACT

A long-running debate in Chinese psycholinguistics has concerned the relative roles of semantics and phonology in reading Chinese characters. Some researchers argue that character reading requires activation of phonological representations, while others maintain the traditional view that Chinese readers generally access meanings directly without phonological mediation. This paper describes an experiment that addresses this debate from a novel direction: Chinese readers were asked to report what they know about simple characters with unfamiliar meanings and/or pronunciations. The "phonology-first" view predicts that it should be impossible to know the meaning of a character without knowing its pronunciation, while the "semantics-first" view predicts that it should be impossible to know the pronunciation of a character without knowing its meaning. Our experiment showed that both situations can exist, though with quantitative and qualitative differences: knowing the pronunciation without knowing the meaning is a somewhat more common occurrence, though it arises most often with characters that share a phonological component with other characters, while knowing the meaning without the pronunciation most often occurs when readers have an alternative (nonstandard) pronunciation for the character or when the character is used as a quasi-linguistic symbol. Moreover, a signal detection analysis found no difference in sensitivity to meaning vs. pronunciation. At the same time, however, readers were strongly biased in their confidence judgments about phonology. Our results thus reaffirm support for a more nuanced position in the debate over Chinese reading, one where both phonology and semantics play key roles in reading.

SUBJECT KEYWORDS
characters, writing systems, reading, phonology, semantics, psycholinguistics

1. INTRODUCTION

A long-running debate in Chinese psycholinguistics has concerned the relative roles of semantics and phonology in reading Chinese characters. Some researchers have argued for a "phonology-first" view whereby all orthographic systems, even that of Chinese, are parasitic on phonological processing (see e.g. DeFrancis 1989, Mattingly 1992, Perfetti and Zhang 1995), while others have maintained the traditional "semantics-first" view that Chinese readers generally access meanings directly from graphic forms without phonological mediation (see e.g. Chao 1968, Hoosain 2002).

A novel way of addressing this debate is suggested by the following two-question quiz which readers may wish to try themselves. First, what is the meaning of the character 廿? Now, how is it pronounced? We don't expect any experienced readers of Chinese to have had any difficulty answering the first question ("twenty"), but those who did have trouble with the second are hardly alone: it is likely that a large proportion of Chinese readers, if not most of them, have trouble recalling the pronunciation of this character (it is, in fact, /nian4/). This is presumably because in spite of the common and readily understandable use of this character in text, it is rarely pronounced. On the face of it, then, the results of this simple test seem to falsify the claim that reading necessarily requires phonological mediation: the meaning of 廿 can apparently be accessed from memory without accessing its pronunciation at all.

In this paper we describe a more sophisticated version of the above test, in which readers were asked to report what they know about simple characters with unfamiliar meanings and/or pronunciations. We begin in Section 2 by providing background on the debate over the roles of semantics and phonology in reading Chinese characters. In Section 3 we describe the design and procedures used in our experiment, spending special care on the selection of our experimental materials (i.e., the unfamiliar characters that we gave to our participants); as we will show, the difficulties faced in the selection procedure are themselves very revealing about the nature of Chinese characters and how they are read. In Section 4 we describe our results, which includes analyses using special statistical techniques for teasing apart actual knowledge about characters from responses biases, and Section 5 summarizes our conclusions.

2. BACKGROUND

Before describing our experiment, we first set it in the context of two interlocking issues. First, are there inherent differences in the accessibility of phonology and semantics from isolated Chinese characters? We suggest that there may well be, with phonology having an advantage over semantics. Second, do readers therefore access the meanings of characters via their pronunciations? The evidence to date strongly suggests that they do not. We then explicate the central question of this paper: Can readers access character meanings without activating pronunciations at all, and/or the reverse?

2.1 The Accessibility of Phonology and Semantics from Isolated Chinese Characters

Extracting the phonological and semantic information from individual Chinese characters is only one small part of the reading process, a process that also involves identifying morphemes and words (see e.g. Taft, Liu, and Zhu 1999; Myers to appear) and their syntactic and semantic roles in discourse context (see e.g. Su 2004). Nevertheless, there is general agreement that the lexical access of characters is a necessary step in this process, and that if we are to understand it for itself, independent of the other complex factors that play a role in reading, the most straightforward approach is to supplement the study of context effects with research on the reading of isolated characters.

The traditional view is that Chinese readers access meanings directly from graphical forms without the mediation of phonological representations (see DeFrancis 1984, 1989 for a historical perspective on this view). While there is much evidence to support this view, as we will see shortly, it is less obvious a claim than naive notions of Chinese characters may suggest. Thus belying the common wisdom that Chinese orthography is an entirely logographic system, DeFrancis (1989) claims that approximately 66% of all Chinese characters contain a phonological radical that is at least partly useful for identifying the character's pronunciation, with approximately 38% of these radicals (25% of all characters) exactly matching the pronunciation of the host character. Perfetti and Tan (1999) provide an even higher estimate, claiming that 85% of all characters have phonological radicals. Moreover, it can be argued that character formation via phonological radical is the most productive form of character coinage, since there is an inverse correlation between the likelihood that a character has this structure and the character's frequency of use; that is, less common characters are more likely to have phonological radicals than more common characters (DeFrancis 1989). Thus while many high-frequency characters are simple (e.g. 人 /ren2/ "person") or are composed of radicals without regular semantic or phonological relationships with other characters (e.g. 能 /neng2/ "be able to"), the opposite is the case with low-frequency characters, which typically have the regular structure of semantic radical plus phonetic radical (e.g. 櫺 /ling2/ "window sill", with semantic radical 木 /mu4/ "wood" and phonological radical 靈 /ling2/ "spirit").[2] This pattern is similar to what has been observed in other lexical systems, such as inflectional morphology (Bybee and Moder 1983), since productivity by definition involves the generation of neologisms, which are generally of low frequency.

Another reason to question the traditional view is that, based on what little we know about differences between semantic and phonological knowledge in general, it seems that semantics is inherently harder, given that semantic representations are much more poorly defined than phonological representations. This can be seen by the great success of universal phonological feature theory (e.g. Chomsky and Halle 1968, and its descendents) as contrasted with the lack of a generally accepted theory of universal semantic features (noted, among other places, by Fodor 1981). It is also reflected in Chinese characters themselves: characters sharing a phonological radical are often homophones, but characters sharing a semantic radical are rarely synonyms. This pattern, however, may arise due to processes that are only relevant during character coinage, or even to the existence of powerful constraints against synonyms (such as the mutual exclusivity bias found in child language; see Markman and Wachtel 1988).

There also seems to be an inherent difference in the accessibility of the phonology and semantics from isolated characters, with phonology again having the advantage. In spite of the engrained habits of dictionary makers, characters are almost never read as isolated lexical items. Most words in Chinese contain more than one character; estimates of the proportion of the vocabulary consisting of two characters range from 50% (Kuo, et al., 2001) to over 70% (Zhou and Marslen-Wilson, 1994). Since a character may represent different morphemes when it appears in different words, seeing it in isolation may make it difficult to come up with a single pronunciation and meaning. However, identifying its meaning is inherently a more difficult task. While there are characters associated with more than one pronunciation, they are relatively rare; Tsai (2000) reports that a mere 4% of characters (540 characters out of a database of 13,060) have two or more pronunciations. By contrast, it seems to be the typical case for meanings and even syntactic categories of characters to depend on the word in which they appear, though it is difficult to quantify this impression since there are no foolproof criteria for distinguishing homonymy, polysemy, and sense extension. Just based on an informal count of dictionary definitions, it is certainly the case that a character is usually given a single pronunciation but multiple, often very divergent, meanings. Even when a character has more than one pronunciation, these pronunciations are typically very similar and much fewer in number than the number of definitions. In *Far East Chinese-English Dictionary* (Liang 1985), for example, the character 累 is associated with three pronunciations that differ only in tone (/lei3/, /lei4/, /lei2/), but it is given a total of eight definitions, some of which have further shades of meaning and changes in syntactic category: /lei3/ can be translated as "to accumulate through a length of time; to pile up; repeat, repeatedly, successive(ly)", /lei4/ as "to involve, involvement, trouble, to implicate; to owe, to be in debt; tired, weary; family burden", and /lei2/ as "a nuisance". Packard (2000) gives several additional examples of characters that shift syntactic category depending on what word they appear in, such as 助 /zhu4/ "help", which is a verb in the coordinative compound 幫助 /bang1zhu4/ "to help", but a noun in the verb-object compound 求助 /qiu2zhu4/ "to seek help". He also points out that semantic shifts in characters due to compound reduction are common as well; an example similar to the ones he cites is 機 /ji1/, which means "machine" in 飛機 /fei1ji1/ "airplane" (literally "fly-machine"), but "airplane" in 機場 /ji1chang3/ "airport" (literally "machine-area").

The above observations may help explain experimental psycholinguistic results suggesting that pronunciations are easier to extract from isolated characters than are meanings. A study by Taft and Zhu (1995) found that the speed for naming characters that always appear in first position of two-character words (e.g. 殉 /xun4/ in 殉國 /xun4guo2/ "die for one's country" and 殉職 /xun4zhi2/ "die while performing one's work") was no faster than the speed for naming characters that only appear in second position (e.g. 侶 /lyu3/ in 伴侶 /ban4lyu3/ "companion" and 情侶 /qing2lyu3/ "sweethearts"). This contrasted with their finding of such a difference for so-called binding characters that only appear in a single word (e.g. 蚯 /qiu1/ was faster to name than 蚓 /yin3/; these two characters only appear in 蚯蚓 /qiu1yin3/

"earthworm"). These results imply that aside from binding words, the pronunciation of individual characters can be accessed without having to access whole word forms first. By contrast, Myers, Derwing, and Libben (2003) found that the semantics of characters were activated more strongly and consistently when presented in compounds than when presented in isolation. They presented Chinese character strings in sequence on a computer screen, with the first string (the prime) serving to prepare the reader for making a decision about the lexicality (i.e. real-word status) of the second item (the target); prime-target pairs were either one-character words as primes and compounds containing them as targets, or the reverse. Response speed to the target was more strongly affected when primes were compounds and targets were single characters than when primes were single characters and targets were compounds. This finding may have more than one explanation, but surely one relevant factor is that character-to-compound priming is weaker than the reverse because isolated characters activate too many competing meanings (see also Tan, Hoosain, and Peng, 1995, for arguments that Chinese readers can use pronunciations of characters even when their meanings are not clear).

2.2 The Roles of Phonology and Semantics in Efficient Chinese Reading

Given arguments like those above, one might expect that Chinese readers routinely access meaning via phonology, in direct contrast to the traditional view. This is in fact the claim of those researchers who argue that all orthographic systems, even Chinese, are parasitic on innate phonological processing, so that reading for reading must necessarily be mediated by phonological representations (DeFrancis, 1989; Mattingly, 1992; Perfetti and Zhang, 1995). It is uncontroversial that the pronunciations of characters can be activated at some point during reading; after all, this is how people read aloud. It also appears that pronunciations may sometimes be obligatorily activated even when overt speech is not necessary. Thus Tzeng, Hung, and Wang (1977) found that the ability to read silently a list of characters in order to memorize them for later recall, as well as the ability to read silently a sentence in order to judge its grammaticality, were both adversely affected if the characters had closely similar pronunciations. However, results like these are consistent with the hypothesis that during certain tasks people habitually use inner speech as a memory aid (what psychologists call the phonological, articulatory, or acoustic loop; see e.g. Zhang and Simon, 1985). It doesn't mean that readers usually read this way, and in fact, Tzeng and Hung (1980) found that when readers had to judge rhyming or to judge sentence grammaticality while simultaneously repeating back a series of digits played over headphones, they showed serious disruptions in the rhyme judgment task but not in the grammaticality judgment task, implying that the latter task did not involve inner speech.

Recently there have been attempts to show that phonological mediation is obligatory in reading Chinese, but these attempts have faced serious empirical challenges from other researchers. For example, Perfetti and Tan (1998) describe a series of priming experiments with single characters as both primes and targets. When the prime was phonologically related to the target, response times for pronouncing the target character were reported as significantly faster relative to targets with control primes or even with semantically related primes. More importantly, by varying the length of time between prime and target, the authors claimed that the phonologically related primes affected response times to the targets slightly earlier than did the semantically related primes. This seems to imply that the pronunciation of a character is activated before its meaning. Unfortunately, in two separate attempts to replicate these results using precisely the same materials and procedures, Chen and Shu (2001) not only failed to find an overall advantage for phonology or evidence that phonology was activated earlier, but in fact found the reverse: semantics was consistently activated earlier and more strongly than phonology. Chen and Peng (2001) and Chen, Wang, and Peng (2003) followed this up with further priming experiments, and never found that phonology is activated earlier than semantics. Similarly, the key findings reported by Tan and Perfetti (1997) in support of the claim that characters can prime characters that are semantically related to their homophones (thus implying that phonology mediates between processing of the graphical form and activation of meaning) could not be replicated by Zhou and Marslen-Wilson (1999).

Evidence for a direct semantic route in reading Chinese characters, and the lack of phonological mediation, may seem to conflict with the productivity of phonological radicals found in the orthographic system itself. What are phonological radicals there for, if not to help readers? In fact, they don't seem to help readers much at all. In contrast to the 26 letters used by English readers, there are over 800 phonological radicals (Hoosain 2002), whose pronunciations must simply be committed to memory. If only about 38% of these exactly match the host character's pronunciation, as noted above, the safest strategy may be for readers not to make use of them for accessing phonological representations (just as one may rationally ignore the reports of a weatherman who has a comparably low accuracy rate). It does seem to be the case that Chinese readers automatically process character components (see e.g. Taft and Zhu, 1997), but these components are not activated solely for their phonological value, since most of them are also characters in their own right and thus are associated with meanings as well. Thus, as Zhou and Marslen-Wilson (2002) found, when readers were asked to judge the semantic relatedness of character pairs, their decision times were slowed down (indicating momentary confusion) if one of the characters was semantically related to the phonological component of the other.

The implication is that reading via phonological radicals should be a very inefficient strategy, and in fact Chen (2002) provides evidence that it is. In an experiment where readers had to judge whether two-character strings were real words (i.e. 詞 /ci2/), the fastest responses were obtained for readers who used a semantics-based strategy (as indicated by the sensitivity of their response times to the concreteness of the word meanings). Slower responses were found for readers who used a phonology-based strategy, specifically those who showed evidence of oversensitivity to phonological radicals (as indicated by their response times to nonsense strings that would sound like real words if one of the characters' components were pronounced as a free character). Similarly, Shen and Bear (2000) found that in the early grades, children learning to write Chinese tended to adopt phonological strategies, as shown by the sound-based errors in their "invented spellings", but as their experience with Chinese orthography increased, they shifted to graphical and semantic strategies. This sort of finding hints at the real reason for the productivity of phonological radicals: inventors of characters, who are faced with the task of transcribing spoken words for the first time, may tend to follow a phonological strategy. Readers face quite different challenges than character coiners, and they do not seem to activate phonology on a regular basis.

Most advocates of the traditional direct-semantics view hold that this is only possible given the special characteristics of Chinese orthography. For example, Hoosain (2002, p. 130) says that the claim that there is direct access to meaning in reading Chinese is "one of the truest myths in Chinese mythology" (playing on an unrelated comment from Chao, 1968, who was, however, also an advocate of this claim), and cites experimental evidence that the access of phonology is slower for readers of Chinese than for readers of English, while the reverse is true for the access of meaning. A variant on this traditional view is that direct access to meaning is the norm for all orthographies, even those that encode phonology relatively transparently. In support of this, experiments by Baluch and Besner (1991, 2001) on Persian, which has an alphabetic orthography, found that the semantic properties of words affected performance on phonological tasks, at least under certain conditions, and Taft and van Graan (1998) report experiments in which English readers were unaffected by phonology when making semantic decisions about written words. Regardless of how the larger theoretical issues are sorted out, it seems reasonable to us to follow Hung and Tzeng (1981, p. 399) and conclude that "phonetic mediation is just one of the strategies for obtaining access to meaning rather than an obligatory stage" (see also Zhou, Shu, Bi, and Shi 1999).

2.3 Phonology without Semantics, Semantics without Phonology

In short, character pronunciations are in principle easier to access than character meanings, and yet efficient readers seem normally to follow a direct route to semantics. Though both sets of observations are well supported, it would be nice to be able to reconcile their apparent contradiction. One way to do this would be to look at the process whereby readers access pronunciations and meanings from isolated characters when they are allowed to ponder them at their leisure, rather than having to make quick decisions as in most of the experiments above (or in fluent reading, for that matter). This sort of study could shed light in two ways. First, use of a novel experimental task could test the assumption, made by Hung and Tzeng (1981) and others, that reading strategies vary according to the reader's needs in a particular situation. Second, a slower task using isolated characters could reveal something about response biases. That is, since phonological representations are inherently clearer than semantic ones, readers should be expected to be more influenced by the phonology of a character when given time to ponder than when they have to make a quick decision about it. Yet this influence should only affect response biases, not the actual amount of knowledge about the character, which would still be accessed primarily for its meaning.

The particular task we chose was a novel, and we hope, illuminating one. In our experiment, we simply presented readers with characters with unfamiliar meanings and/or pronunciations. If reading primarily involves activation of phonology, it should be impossible (or at least very difficult) to know the meaning of a character without knowing its pronunciation, whereas if reading primarily involves activation of semantics, it should be impossible (or at least very difficult) to know the pronunciation of a character without knowing its meaning.

Our initial hunch was that the latter situation should be much more likely. This intuition was based on a small collection of simple characters that seemed to us to be used confidently by fluent readers and writers for their meaning, without the users seeming to know how they were pronounced; the fact that we had never observed the opposite situation was a primary motivation for this study. These characters were 廿, which, as noted in the introduction, all literate Chinese know means "twenty" though many are unaware of its pronunciation /nian4/; 卅 "thirty", for which even fewer know the pronunciation /sa4/; 卍, the familiar Buddhist symbol with the unfamiliar pronunciation /wan4/; and the two characters in 孑孓 /jie2jue2/ "mosquito larva". The fact that 孓 appears in only this one word (孑 also appears in the idiom 孑然一身 /jie2ran2yi1shen1/ "alone") makes it particularly interesting to examine, since if even a binding character can be recognized for meaning without accessing its phonology, this would provide particularly strong evidence for a direct route between orthography and semantics, with no mediation by phonology even at the word level.

We now turn to a description of how we designed and conducted this experiment.

3. THE EXPERIMENT

The questions we wanted to address in our experiment were quite simple: Is it possible to be aware of the meaning of a character without being aware of the pronunciation, and is it possible to be aware of the pronunciation of a character without being aware of the meaning? However, a variety of conceptual and methodological challenges had to be faced if such an experiment was to provide interpretable results.

Most fundamentally, as an experiment involving a number of items and participants, our experiment cannot be expected to provide a sharply categorical answer (i.e., possible vs. impossible). Since a particular response may indeed prove to be possible, but so rare that its occurrence can't be distinguished from chance (e.g., some participant just happened to make a lucky guess), we must analyze the results quantitatively using standard statistical procedures.

Quantitative analysis, however, raises a problem of its own. Inferential statistics, which allow one to generalize from a sample of observations (e.g. the characters we chose for our experiment) to an entire population (e.g. all Chinese characters), typically assumes that the sample is chosen randomly from the population. This is almost never true in actual psycholinguistic practice; experimental participants are usually selected arbitrarily, but not truly randomly, and linguistic materials are intentionally pre-screened very carefully to control for irrelevant variables expected to skew the results. For example, a study testing whether function words are processed more quickly than content words had better control for lexical frequency as well, since this variable is known to exert an independent effect on processing time; however, since most function words are more frequent than most content words, results using a well-controlled set won't actually generalize to the lexicon as a whole. Usually this sort of situation is not considered a problem, since the more immediate goal is not to predict the behavior of words in general, only the behavior of words in further experiments, which will require items to be controlled in basically the same way. Similarly, for the reasons described below, we must also select the characters for our experiment very carefully. In our case, however, we would indeed like to know something about what our experiment means for reading Chinese characters in general, so the lack of randomness in our sample means that we must be very cautious in drawing wider conclusions. Ironically, this caution requires that qualitative observations about individual items will have to play a greater role in our analysis than is typical in psycholinguistic research.

In our case, the most important irrelevant factor that we must control is the relationship between a character used in the experiment and other characters that the participants might know; we cannot allow participants to be able to guess the meaning and/or pronunciation of the experimental character using anything but knowledge of this character alone. The most obvious consequence is that our items cannot be complex, that is, composed of radicals that are themselves characters or are used as radicals in other characters. Thus we could not use the relatively unfamiliar character like 騮 /liu2/ "a legendary horse", since something about its meaning can be guessed from its semantic radical (馬 /ma3/ "horse") and something about its pronunciation can be guessed from its phonological radical (留 /liu2/ "remain"). A similar constraint applies to simple characters that are used as radicals in more complex characters. Thus we should not use 豕 /shi3/ "pig", due to its appearance in semantically related characters like 豬 /zhu1/ "pig", and we should not use 囪 /cong1/ "chimney", due to its appearance in phonologically related characters like 聰 /cong1/ "clever". In practice, these criteria proved difficult to apply strictly. Again this underlines the necessity of supplementing overall statistical analyses with discussion of individual items.

A final challenge arises from the fundamental differences between semantics and phonology discussed in the previous section. If semantic representations are inherently vaguer than phonological representations, it doesn't seem fair to pit one against the other without some correcting adjustments. In particular, even if character reading does in fact primarily rely on semantic processing it would not be surprising to find that our experimental participants show a strong response bias regarding confidence about phonological knowledge. The primary way we dealt with this methodological challenge was to employ signal detection theory, a mathematical model of perception that allows one, in principle, to distinguish response bias from inherent sensitivity to a stimulus input (here, the participants' memory representations of character meanings and pronunciations). Details concerning our application of signal detection theory will be described in Section 4 when we discuss the results.

3.1 Selection of Characters

Our experiment was intended to challenge the character access system of a fluent reader with extreme cases, in order to reveal something about the inherent biases and limitations of this system. Thus rather than selecting a representative sample from the set of all Chinese characters, we wanted characters of a very particular sort, namely, those likely to be unfamiliar in either pronunciation or meaning, yet without containing orthographic clues to pronunciation or meaning. These selection criteria are relatively clear, but implementing them proved to be quite difficult. In the end we used three sources: our original small collection of "partially familiar" characters, a free character production pretest, and character frequency databases. Although it is debatable (as a reviewer points out) that these sources are truly optimal, they seem intuitively reasonable: start with characters that linguists consider to be theoretically interesting, ask nonlinguist readers to suggest further examples, and finally add characters of sufficiently low frequency that many readers will be less than fully familiar with them.

Our initial small set consisted of characters that seemed to us to have familiar meanings but unfamiliar pronunciations, namely 廿, 卅, 卍, 孑, 孓. This set is hardly unbiased, of course, and each of our characters raise its own difficulties as well. Thus 孑 and 孓, though simple characters, are related in both form and meaning to 子 /zi3/ "child, seed". The characters of 廿 and 卅 are somewhat iconic, representing visual blends of 十 /shi2/ "ten" with 二 /er4/ "two" and with 三 /san1/ "three", respectively. It is also a bit misleading to claim that their pronunciations are unfamiliar. True, their monosyllabic dictionary pronunciations may be unfamiliar, but every Chinese reader will readily associate 廿 with the phonological representation /er4shi2/, which is the pronunciation for the two-character equivalent 二十 ("twenty"), and 卅 with the phonological representation /san1shi2/ (三十 "thirty"). Indeed, these are the standard pronunciations used when reading aloud passages containing these characters, just as 二個 /er4ge0/ "two (things)" is read aloud as /liang3ge0/ (i.e.,

兩個, using the allomorph for "two" appropriate for a classifier construction), and some speakers in Taiwan pronounce the character 元 /yuan2/ "dollar" as /kuai4qian2/ (i.e., 塊錢 "unit of money"). Finally, the character 卍 is highly familiar as a prominent Buddhist icon, but it's not so clear if it can be considered a true linguistic constituent, such as a morpheme or word. Dictionaries do list it as a noun, but a search for it on Google.com suggest that it is most commonly used as a nonlinguistic symbol: in 94% of the hits for 卍, it was immediately followed by 字 /zi4/ "character" or 號 /hao4/ "symbol", and in 99% of these, the character appeared within quotation marks. The symbol is also often used in decorations (3% of the Google.com hits involved repeated strings of 卍). Even if it is a word, it may be unfair to ask anyone not steeped in Buddhist theory what its precise meaning is. Moreover, like 廿 and 卅, for many readers it is also associated with a multisyllabic (and multimorphemic) phonological representation, namely /fo2jiao4/ (佛教 "Buddhism").

We thus decided to supplement this small set with characters from two other sources that we hoped would be more objective: a free character production pretest and character frequency databases.

3.1.1 Free Character Production Pretest

In the free character production pretest, thirty college students studying in southern Taiwan were simply asked to list on a paper form as many characters as possible for which they felt they did not know the meaning and/or the pronunciation. For each character, they were also asked to provide whatever knowledge they did have about it, leaving the associated space blank if they felt that they knew nothing. To reduce any bias caused by the instructions themselves, the order of the words "meaning" and "pronunciation" was counterbalanced across participants (i.e., half received instructions with the above order, and the other half instructions with "pronunciation" before "meaning"). The two Chinese versions of the instructions that we used are shown in Table 1.

Table 1. Instructions for the pretest

請寫出你不知道意義或是不知道發音的字。請在第一欄中寫出字，第二欄中寫出這個字的發音或是意義。（若都不知道，第二欄可不填。）

請寫出你不知道發音或是不知道意義的字。請在第一欄中寫出字，第二欄中寫出這個字的意義或是發音。（若都不知道，第二欄可不填。）

This pretest gave us 288 characters which were self-reported to be at least partly unfamiliar to at least one of these 30 students. Fifteen of these characters, however, were so poorly remembered that even the form written down by participants was unrecognizable; these were eliminated. Based on the accuracy of the semantic and phonological descriptions given by the participants, as judged by the third author, we then identified 187 characters which at least one of the participants had described accurately in its semantics and/or phonology. For the vast majority of these 187 characters, at least one participant had correctly described only the phonology (172/187 characters, or 92%), dwarfing the number of characters for which at least one participant had correctly described the semantics (9/187 characters, or 5%) or for which both phonology and semantics were correctly described by at least one participant (6/187 characters, or 3%). By itself, this is not very meaningful for our purposes, given that virtually all of these characters were complex, and many had transparent phonological radicals. At the same time, while all five of the characters in our original collection were spontaneously offered by at least one participant in the pretest, only 廿 was given a correct semantic description by the one participant who listed it (without a pronunciation). The one participant who listed 孑 and 孓 gave no pronunciations and only partially correct definitions, and 卅 and 卍 were given correct pronunciations by one participant each, who nevertheless failed to list any meanings (though surely they knew them).

We then focused our attention on the 73 characters which no participant had described accurately in either pronunciation or meaning. This was done both to maximize the likelihood that participants in the main experiment would be unfamiliar with the characters that we would give them, and also to reduce any inherent bias in the materials towards phonology or towards semantics. However, most of these 73 characters also failed to meet basic criteria for use in the main experiment.

First, virtually all of them were complex, with clearly identifiable radicals; depending on how generous one was about semantic and phonological relatedness, up to 65 of the 73 characters contained radicals related in meaning or pronunciation with other characters. We did not want to use any character of this sort, for the reasons explained above. Some of these complex characters contained clues to both semantics and phonology, such as 詁 /gu3/ "to perform an exegesis". In spite of its low frequency, it is formed in an entirely regular way: the semantic radical 言 /yan2/ "speech" indicates its relationship with language, and the phonological radical 古 /gu3/ "ancient" indicates its pronunciation. Thus even for a reader unfamiliar with this character it would be easy to guess something about it simply by familiarity with other characters. Another example like this was 輿 /yu2/ "carriage", with its semantic radical 車 /che1/ "vehicle" and phonological radical 與 /yu3/ "and". Other characters contained clues only to their semantics, such as 嬲 /niao3/ "to flirt with", composed of 男 /nan2/ "male" and 女 /nyu3/ "female", while still others only contained clues to their phonology, such as 浡 /bo2/ "rise, excited", composed of the reduced form of the radical 水 /shui3/ "water" and the phonological radical 孛 /bo2/ "change facial color". In many characters, however, the relationships with other characters, while present, seemed to be weaker. For example, in the character 毗 /pi2/ "to assist, adjoin", the phonological radical 比 /bi3/

"compare" has a pronunciation similar but not identical to that of the character as a whole, but this radical also appears in other characters that do have an identical pronunciation with that of 毗, such as 琵, the first character in 琵琶 /pi2pa2/ "Chinese lute". Naturally, semantic relationships with other characters were even more unclear, due to the inherently vague semantics of the radicals themselves. For example, in the suggested character 鉉 /xuan4/ "device for carrying a tripod", the 金 /jin1/ "gold" radical indicates the class "device", while in the suggested character 鋰 /li3/ "lithium" it indicates the class "chemical element". Even where the semantic radical indicates an unambiguous aspect of character's meaning, it is impossible to reconstruct all of the meaning from this radical alone. Thus although both the suggested character 耄 /mao4/ "in extremely old age" and the suggested character 耋 /die2/ "in one's eighties" are semantically related to the radical 老 /lao3/ "old", this radical isn't sufficient to identify what makes the meanings of these characters unique. By contrast, note that the phonological radical of the first of these two characters (毛 /mao2/ "hair, fur") exactly matches the pronunciation of the host character (the phonological radical in 耋, namely 至 /zhi4/ "to", is admittedly much less helpful to readers, being associated with a pronunciation similar to /die2/ only in characters even rarer than 耋).

A second problem with several of the characters contributed in the free character production pretest was that their meanings were inherently unclear in modern Chinese, being used today primarily as proper names, as transliterated foreign borrowings, or even as parts of function words. An example of such a character suggested in the pretest was 湛 /zhan4/. Although it is associated with the meanings "dewy" and "deep", it is more familiar as a surname. An example of a character used today solely in transliterations is 伽, though in one of those bits of inefficiency that make Chinese orthography so charming for learners, it is associated with two distinct (albeit similar) pronunciations: /jia1/, as in 伽馬 /jia1ma3/ "(Vasco) da Gama", and /qie2/, as in 薄伽尼 /bo2qie2ni2/ "(Umberto)

Boccioni". Finally, the character 麼 /me0/ was suggested in the pretest; even though it appears in the high-frequency function words 什麼 /she2me0/ "what" and 怎麼 /ze3me0/ "how", it is indeed unclear what it itself means. Yet it seems unlikely that an educated reader of Chinese would fail to know all that one needs to know about 湛, 伽, and 麼 in order to use them properly, so it is not at all obvious that the existence of such characters implies that semantics can be bypassed in reading. Thus it seems unfair to make a majority of our experimental items be characters of this sort, since they are biased against semantics. At the same time, however, since we began our collection with examples that are biased against phonology, we had no justification for keeping them out entirely. In any event, the three examples just cited had to be ruled out anyway, since they are complex, containing clear radicals and phonological radicals of varying degrees of usefulness.

In the end, we only chose two characters from the free character production pretest to add to our original collection, though neither is entirely bias-free. If one were unfamiliar with the relatively low-frequency character 勺 /shao2/ "spoon", it is in principle possible to guess its pronunciation or meaning from characters that use it as a phonological radical, such as 杓 /shao2/ "spoon" and 芍 /shao2/ "shao (type of plant)", but these are themselves quite unfamiliar characters. The character 仝 /tong2/ "same" is also likely to be quite unfamiliar to many readers, and it is not used as a radical in other characters. However, some readers may know that it is an alternative form for the very familiar character 同 /tong2/ "same", and it may also be familiar as a surname.

3.1.2 Character Frequency Databases

Given the problems noted above, we decided to supplement our small collection with characters chosen from large corpora, which have a wider selection of low-frequency (hence less familiar) characters. The primary corpora that we used were the Ministry of Education (MOE) corpus of approximately 2,000,000 characters (Li, Li, and Tseng 1997), based on educational materials used in Taiwan, and the Academia Sinica Balanced Corpus of approximately 8,000,000 characters (Chen, Huang, Chang, and Hsu, 1996), based on a variety of materials written in traditional Chinese characters. We also supplemented these sources with searches on Google.com, which allow one to estimate token frequency by counting number of Web page hits (see Blair, Urland, and Ma, 2002, for empirical justification of the validity of using Web search engines to estimate frequencies).

Of course, simply identifying the lowest-frequency characters was not enough for our purposes, since as noted earlier, most low-frequency characters contain phonological radicals. We thus restricted our search to characters which were not only below a fixed frequency threshold (arbitrarily set to 50 tokens in the MOE corpus), but which also had fewer than a fixed number of strokes, as a crude measure of simplicity (arbitrarily set to 8 strokes). In our attempt to purge our experimental items of characters related to other characters, we first rejected any character with a semantic radical in its standard position, even if its semantic contribution wasn't clear (e.g. 仆 /pu1/ "fall forward", with the radical 人 in its reduced form on the left). It was impossible to be entirely strict about radicals, however, since most simple characters that don't properly contain a radical are themselves radicals, and so had much too clear a relationship with other characters to use in the experiment. Thus we did accept some characters with clearly delineated radicals, although we did attempt to make it difficult for readers to know what the radicals were (e.g. 仄 /ze4/ "slanting, oblique", the lexicographic radical in which is 人, not 厂).

Once we had settled on a small set of low-frequency, apparently simple, non-radical characters, we were reluctant to reject some merely because they also formed radicals in other characters, since to do so would have whittled our choices down to very few indeed. In the case of our character choices, several involved phonological radicals, though one semantic radical was involved as well. Table 2 lists these ten potentially problematic characters, along with the characters in which they appear as radicals. Meanings and pronunciations are only given for the one semantically related character; for the rest, the meanings are irrelevant and the pronunciations are identical to that of the target character. Phonological radicals that appear in characters with related but not identical pronunciations are not listed here. Since these related characters can help readers in our experiment only if they are more familiar with them than with the target characters themselves, the table also gives the frequencies of these other characters in relation to the given target character's frequency: values of the ratio frequency(related character) / frequency(target character) over 1 indicate situations where a related character is likely to be more familiar than the target character (bolded in the table). Not all of these characters are found in our two primary corpora, so here frequencies were estimated using the number of Web page hits for each character in a search on Google.com (conducted on March 4, 2004) of Web pages with traditional Chinese characters (Big5 encoding).

Table 2. Characters in our materials that also serve as radicals

Target characters	Related characters	Relative frequency*
亍 /chu4/ "a step with the right foot"	行 /xing2/ "walk"	**54.91**

[Table continued on next page.]

Table 2. [continued]

Target characters	Related characters	Relative frequency*
壬 /ren2/ "ninth of ten celestial stems; artful; great; (counting symbol)"	任**	**15.09**
	銋	0.86

	紝	0.80
乍 /zha4/ "unexpectedly"	炸	**1.90**
	咋	0.87
	詐	0.80
	痄	0.67
	鮓	0.66
	搾	0.60
	蚱	0.60
	砟	0.59
	醡	0.59
囪 /cong1/ "chimney"	聰	**1.74**
	蔥	**1.68**
	驄	**1.05**
	璁	**1.04**
	蟌	0.97
	熜	0.84
卞 /bian4/ "hurriedly; (proper name)"	汴	**1.10**
	抃	**1.02**
	忭	0.92
	苄	0.92

Table 2. [continued]

Target characters	Related characters	Relative frequency*
兀 /wu4/ "cut off the feet; high; this; etc"	靰	0.94
	阢	0.89
	杌	0.84

冉 /ran3/ "tender"	玶	0.91
	姌	0.85
	呥	0.76
	髯	0.75
卮 /zhi1/ "wine container"	栀	0.90
叵 /po3/ "cannot"	笸	0.87
亙 /geng4/ "extreme"	埂	0.75

* freq(related)/freq(target)

**Though this character is very frequent, it is almost always pronounced /ren4/, not /ren2/.

Note that the character 卮 apparently never serves as a phonological radical in precisely this form, but a graphically similar form appears in other characters with the same pronunciation (e.g. 栀 /zhi1/ "gardenia"). Some of the other characters listed above also have special properties that might give a participant in our experiment an unfair advantage in guessing the pronunciation, meaning, or both. Thus the character 卞 is most familiar as a proper name. Similarly, the character 壬 may not be familiar in terms of its classical meaning "ninth of ten celestial stems", but it may be familiar as an arbitrary counting symbol, ninth in the series beginning with 甲 /jia3/, 乙 /yi3/, and 丙 /bing3/ (though this series is rarely carried out so far; a Google search found 甲 to be almost 14 times more frequently used than 壬). The character 叵 appears in the idiom 居心叵测 /ju1xin1po3ce4/ "can't tell someone's true intentions", which may make its pronunciation more familiar than its meaning. Finally, our inclusion of the character 囱 requires a brief note. In addition to being a common phonological radical, this character also appears in the familiar compound 煙囪 /yan1cong1/ "chimney". The only properties that may lead readers to have difficulties with this character are its relatively low frequency and its restriction to this single compound.

In addition to these characters, our filtering of the information from the corpora resulted in the selection of seven additional characters without any clear semantic or phonological relationships with other characters. These were 仄 /ze4/ "slanting, oblique" (a technical term from poetry and music); 乜 / mie1/ "glance sideways" or /mie4/ "(surname)"; 弁 /bian4/ "low-ranking military officers"; 朮 /zhu2/ "kind of plant"; 丫 /ya1/ "fork, crotch"; 囟 /xin4/ "top of the skull" (especially the undeveloped part of a baby's skull); 万 /wan/ "ten thousand". As is true for the rest of our collection, none are entirely free of problematic features. Thus 仄 is technically a complex character, as described above, and so is 弁 (with the lexicographic radical 廾); 乜 has two different pronunciations (making its phonology less clear) and is used as a surname (making its semantics less clear);丫 is both highly iconic and identical in form to the phonemic zhuyin fuhao (注音符號) symbol ㄚ pronounced /a1/; the meanings of 朮 and 囟, though concrete, are relatively technical and thus may be difficult to explain; and 万 is an alternate form of 萬 /wan4/ "ten thousand" (though 万 is not the standard way of writing this morpheme in Taiwan, some participants may know that it is the standard form in simplified orthography).

3.1.3 Final Set of Experimental Materials

The set of 24 characters resulting from our selection procedures is given in Table 3. The characters are listed with their pronunciations, meanings as given in standard Chinese-English dictionaries (Liang 1985, *General Chinese-English Dictionary* 1997), source (original set, free production pretest, or corpora), and frequency in number of tokens in our two corpora.

Table 3. Materials used in experiment

				Token frequency	
		Gloss	Source	Sinica*	MOE**
卅	sa4	Thirty.	Original /Pretest	0	0
孑	jie2	Halberd. Only; single. Solitary. Alone.	Original /Pretest	0	0
孓	jue2	Only used in 孑孓 jie2-jue2: Mosquito larva	Original /Pretest	0	0
卍	wan4	Swastika. A sign of Buddhism. Ten thousand.	Original /Pretest	0	1
卮	zhi1	Container for wine.	Corpora	0	1
朮	zhu2	Name of a plant.	Corpora	0	1
仝	tong2	Same; equal; similar; identical. Agreeing. All; united.	Pretest	0	1

[Table continued on next page.]

Table 3. [continued]

				Token frequency	
		Gloss	Source	Sinica*	MOE**
万	wan4	(n.) Ten thousand. (adj.) Myriad; multitudinous. (adv.) Absolutely;	Corpora	2	0

		extremely; by all means.			
叵	po3	Cannot.	Corpora	0	4
仄	ze4	An oblique tone; consonant; inclined. Inclined; slant.	Corpora	0	5
弁	bian4	A military cap; a conical cap. Military officers of a low rank. Preface; foreword. Alarmed; frightened. Quickly; hurriedly	Corpora	0	6
亍	chu4	A step with the right foot.	Corpora	8	0
卞	bian4	Hurriedly; rash. Excitable A family name. Kaifeng (capital of Song Dynasty)	Corpora	9	0

[Table continued on next page.]

Table 3. [continued]

				Token frequency	
		Gloss	Source	Sinica*	MOE**
乜	mie1/ mie4	mie1: Glance sideways. mie4: A family name.	Corpora	13	1

廿	nian4	Twenty; twentieth.	Original /Pretest	8	8
勺	shao2	A spoon; a ladle. A small quantity; a little bit. Take with a spoon.	Pretest	11	7
亙	geng4	The first or last quarter of the moon. An extreme limit. Universal. Fill or extend.	Corpora	12	13
囱	cong1	A chimney; a flue.	Corpora	15	13
兀	wu4	Cut off the feet flat. High and on the top. Ignorant-looking. This.	Corpora	16	16
囟	xin4	Front part of the skull of a baby that is not yet fully developed.	Corpora	34	2

[Table continued on next page.]

Table 3. [continued]

				Token frequency	
		Gloss	Source	Sinica*	MOE**
壬	ren2	The ninth of the Ten Celestial Stems. Artful and crafty. Great. Pregnant.	Corpora	36	8

丫	ya1	A fork. A crotch.	Corpora	9	43
冉	ran3	Tender; weak. Gradually.	Corpora	61	14
乍	zha4	Unexpectedly; suddenly; abruptly. At first; for the first.	Corpora	71	41

*Academia Sinica Balanced Corpus

**Ministry of Education corpus

3.2 Procedure

3.2.1 Participants

Twenty college students in southern Taiwan (native speakers of Mandarin and fluent readers of traditional Chinese characters) were paid for their participation in the experiment.

3.2.2 Task

Participants were given a sheet of paper with a table in which each of the 24 characters (no filler items were used) appeared on a separate line, with a row of four spaces for participants to write in their guesses about pronunciations and meanings, as well as their degree of confidence in each of these guesses. At the top of the sheet were the instructions shown in Table 4.

Table 4. Instructions for main experiment.

請寫出下列各字的發音(注音符號)和意義(你知道有關這個字的任何事)及你對此答案的確定性，非常不確定請填1；非常確定請填5。

<u>不要</u>留下空白，如果只是猜測，請在「確定性」欄填1，謝謝。

不確定........... ....確定

1 2 3 4 5

[Translation:
Please write down the pronunciation of this character (zhuyin fuhao) and its meaning (anything you know about this character), and also your certainty about your answer: if very uncertain, fill in "1", if very certain, fill in "5". **Do not** leave any space blank; if you're just guessing, fill in "1". Thank you.

Uncertain..........................Certain

1 2 3 4 5]

Due to an oversight, there was only one version of the instructions (that shown above), and this mentioned pronunciation before meaning. Though this may bias the results in favor of phonology, note that the instructions imply rather loose criteria for semantics, namely "anything you know about this character". In fact, in our scoring, mere identification of a character as a proper name was sufficient to count as a correct answer for meaning. Moreover, the arrangement of the spaces for writing down their answers was counterbalanced across participants: half received a version with the space for pronunciation to the left of the space for meaning, and the other half received a version with the reverse order (readers presumably filled in the space on the left before filling in the space on the right). Each of the versions of the list themselves came in two versions depending on the sequence of the characters in the list: half had the items in random sequence, and the other half had the reverse sequence of this. Participants were given no time limit for finishing the form.

Since this task involves the reading of characters in isolation (including 孑 and 孓), without any time limit, it is not a realistic model of Chinese reading in general, where characters are virtually always seen in context, and usually only for very brief periods of time. As mentioned above, however, this experiment is intended to study just one aspect of the reading process, namely the lexical access of individual characters, under extreme conditions, in order to gain a clearer understanding of its inherent biases and limitations. For example, our methods are sufficient to falsify the claim that character pronunciations are obligatorily accessed prior to character meanings, namely if we found that readers know the meanings of some characters without knowing their pronunciations. This would provide an existence proof showing that the character processing component of Chinese reading does not need to rely on phonology.

4. RESULTS AND DISCUSSION

4.1 Coding of Responses

As instructed, all participants gave responses for both pronunciations and meanings for all of the characters. However, there were 15 missing responses for confidence ratings, 10 of which came from one participant (3% out of a total of 480 = 24 items × 20 participants). These were removed from further analyses, though it is important to note that all but one of them were associated with incorrect guesses.

Responses for both pronunciation and meaning were coded into three-point scales, with 0 being the most incorrect and 1 being the most correct. For pronunciation, a score of 1 was given only if the response was perfectly correct; a score of 0.5 was given if the response differed from the correct pronunciation in only one zhuyin fuhao symbol (i.e., initial, medial, rime, or tone); a score of 0 was given if the response contained two or more errors. For meaning, a score of 1 was given if the response was deemed by the third author to be a match with any dictionary definition for the character; this included simply stating that the character was a person's name or a place name. A score of 0.5 was given if the response was a phrase or word that contained the character, rather than defining it, or if the response gave a definition that gave specific information differing partly from the character's dictionary definitions. A score of 0 was given if the response did not seem to be related at all to the dictionary definitions. Examples of scoring are shown in Tables 5 and 6.

Table 5. Scoring of responses for pronunciation.

	Responses for pronunciation		
Character	Scored as 1 (most correct)	Scored as 0.5	Scored as 0 (least correct)
廿	ㄋㄧㄢ 4 (/nian4/)	ㄋㄧㄢ 3 (/nian3/)	ㄍㄢ 1 (/gan1/) ㄦ 4 (/er4/) ㄚ 4 (/a4/)
丫	ㄧㄚ 1 (/ya1/)	ㄚ (/a1/)	(no examples)

Table 6. Scoring of responses for meaning.

	Responses for meaning		
Character	Scored as 1 (most correct)	Scored as 0.5	Scored as 0 (least correct)
乍	突然 ("abruptly")	乍看之下 ("at first glance") 乍現 ("gleam")	植物的名稱 ("plant name") 傾倒 ("dump") 狹小狀 ("narrow and small")

[Table continued on next page.]

Table 6. [continued]

	Responses for meaning		
Character	Scored as 1 (most correct)	Scored as 0.5	Scored as 0 (least correct)
廿	二十 ("twenty")	數字 ("a	唸("read")

		number") 數字，表二 ("a number related to two")	
万	萬的簡字 ("simplified character of 萬") 數字 10000 ("a number for ten thousand") 數字單位 ("a measurement of number")	(no examples)	金錢 ("money") 瓦片 ("a tile")

As a reviewer points out, the attempt by this scoring system to put phonology and semantics on the same scale seems problematic. In particular, we have emphasized throughout the above discussion that semantic representations are inherently vaguer than phonological representations. Since the goal of this study is to understand the roles of phonology and semantics in character reading, separate from any inherent differences in the two types of information, we must face this problem head on. We have adopted a three-pronged approach. First, the basic scoring system was intentionally designed to be more forgiving of semantic mismatches than phonological ones, which is appropriate given the inherent vagueness of character semantics. Thus a score of 0 could be received for pronunciations that matched in as many as two subsyllabic elements (if this involves mismatching in two others), whereas a score of 1 could be received for "meaning" responses even if the only thing known was that the character represented some proper name. Second, as described in section 4.2 below, we also reanalyzed the results using two additional coding systems which may be argued to be even fairer. Third, and most important, in sections 4.3 and 4.4 we describe analyses of the confidence scores, which not only provide a different source of data about how the characters were processed, but also allow us to use signal detection theory to reveal patterns in sensitivity to pronunciations or meanings, separate from any inherent biases.

4.2 Accuracy Scores

We began by conducting standard by-participant and by-item analyses on the raw accuracy scores. In the by-participant analysis, averages were first calculated across all 24 items for scores divided into four categories according to two binary factors: the type of knowledge that was probed (pronunciation vs. meaning) and the order in which these knowledge types were asked for first on the form (pronunciation first vs. meaning first). Knowledge type was a within-group factor (each participant was probed on both types) while order was a between-group factor (each order was given to half of the participants). The analysis thus allowed us to determine not only whether one or both factors had a significant effect on accuracy, but also whether the two factors interacted with each other (e.g. if phonological accuracy was only higher if phonological knowledge was probed first).

The average scores were put into a two-way analysis of variance (ANOVA), which uses F distributions (based on ratios of "interesting" variance to "nuisance" variance) to generate p values, which represent the probability that the results could have been due to chance; following convention, we considered p values below 0.05 to indicate statistical significance. The by-participant ANOVA found that only the factor of knowledge type showed a significant difference ($F(1,18) = 49.37$, $p < 0.0001$): the mean pronunciation accuracy score was 0.59, while the mean meaning accuracy score was significantly lower, at 0.46. Overall accuracy scores were slightly higher for participants who were probed for meaning knowledge first (0.53 vs. 0.52), but this tiny difference was not at all statistically significant ($F(1,18) = 0.025$, $p > 0.87$). There was also no significant interaction between the two factors ($F(1,18)$

= 0.27, $p > 0.60$): regardless of which type of knowledge they were probed for first, participants were consistently more accurate with phonological knowledge.

We also performed a similar analysis by item (character), where we first averaged across participant scores in the same four categories. Since each item thus provided four scores, we treated both knowledge type and order as within-group factors. The resulting two-way ANOVA showed the same pattern as described above: a significant effect of knowledge type ($F(1,23) = 5.96$, $p < 0.023$), but no effect of order ($F(1,23) = 0.28$, $p > 0.60$) and no interaction between the two factors ($F(1,23) = 0.18$, $p > 0.67$). The problem with interpreting the by-item analysis, however, is that unlike our arbitrary set of participants, our items do not represent anything like a random sample of Chinese characters; the ANOVA, like most statistical models, assumes that samples are randomly chosen from the larger population.

A more serious problem with these results is that they are based on two different scoring systems for pronunciations and meanings, as noted in the previous section. Thus the significant effect of knowledge type may be caused by how we scored, rather than any real difference in knowledge. To test this, we ran new ANOVAs on two alternative ways of scoring. In the first, which we'll call the "strict" scoring system, we counted only a perfect score as a correct response; all other responses, even if partly correct, were counted as errors. That is, we recoded both 0 and 0.5 in the original system as 0. This strict scoring may be more objective, since perfectly correct answers were easier for us to identify than partly correct answers. The results of the ANOVAs conformed closely to the original analyses. Mean accuracy scores for pronunciations (0.53) were significantly higher than for meanings (0.31), both by participant ($F(1,18) = 49.91$, $p < 0.0001$) and by item ($F(1,23) = 6.25$, $p = 0.02$), and again there was no significant effect of order and no interaction (Fs < 0.85, ps > 0.77). Thus even with the strict coding system, pronunciation accuracy was higher.

This alternative scoring system may still be biased against semantics, however. Correct answers for pronunciations are inherently easier to identify than correct answers for meanings; the former will exactly match a dictionary entry, while the latter may use an equivalent paraphrase. Moreover, several of our experimental items were phonological radicals while none were semantic radicals. Thus we decided to bias our scoring against phonology to see what would happen. We did this by using the recoding described above, where a raw score of 0.5 (partly correct) was recoded as 0 (entirely wrong), but for phonology only, while for semantics we used the opposite recoding: a raw score of 0.5 was recoded as 1 (for both, an original 1 was coded as 1 and an original 0 as 0). This new recoding thus inflates the accuracy for semantics while deflating that for phonology. Note that it also counts as correct semantic responses in which participants merely gave words or phrases containing the target character; given that the meaning of a character is at least partly dependent on the word in which it appears, this may be the fairest coding system of the three. In any event, with this "semantics-favoring" scoring, the result was that the ANOVAs found no significant effects at all, not even for knowledge type (Fs < 0.47, ps > 0.50).

It may reasonably be objected, however, that regardless of the scoring system, the above analyses generally favor phonology because of the particular set of characters we used as experimental items. If most of these happen to have easier-to-guess pronunciations than meanings, then analyses treating them as a group will naturally show a bias towards pronunciations. Thus we also performed by-participant analyses for each individual character. To simplify the analyses without affecting statistical power, we ignored the effect of order and used paired two-tailed t tests (equivalent to ANOVAs when comparing only two sets) to look at the effect of knowledge type. The resulting differences in mean accuracy scores (mean phonology score - mean semantics score) according to the three scoring systems are summarized in Table 7.

Table 7. Differences in mean accuracy scores (phonology - semantics) according to three scoring systems

	Original			Strict			Semantics-favoring		
	Phon	Sem	**Dif**	Phon	Sem	**Dif**	Phon	Sem	**Dif**
廿	0.38	0.85	**-0.48****	0.35	0.80	**-0.45****	0.35	0.90	**-0.55*****
卅	0.65	0.90	**-0.25****	0.65	0.80	**-0.15**	0.65	1.00	**-0.35****
卍	0.50	0.73	**-0.23**	0.50	0.60	**-0.1**	0.50	0.85	**-0.35***
乍	0.73	0.80	**-0.08**	0.70	0.75	**-0.05**	0.70	0.85	**-0.15**
孑	0.73	0.78	**-0.05**	0.50	0.75	**-0.25**	0.50	0.80	**-0.30***
仄	0.55	0.60	**-0.05**	0.25	0.50	**-0.25**	0.25	0.70	**-0.45****
囟	0.10	0.10	**0.00**	0.10	0.10	**0**	0.10	0.10	**0.00**
仝	0.13	0.10	**0.03**	0.10	0.10	**0**	0.10	0.10	**0.00**
乜	0.18	0.15	**0.03**	0.15	0.10	**0.05**	0.15	0.20	**-0.05**
卮	0.30	0.20	**0.10**	0.30	0.10	**0.20***	0.30	0.30	**0.00**
勺	0.60	0.50	**0.10**	0.60	0.50	**0.1**	0.60	0.50	**0.10**
万	0.88	0.75	**0.13**	0.85	0.75	**0.1**	0.85	0.75	**0.10**
亘	0.40	0.28	**0.13**	0.20	0.05	**0.15**	0.20	0.50	**-0.30***
孓	0.90	0.75	**0.15**	0.80	0.70	**0.1**	0.80	0.80	**0.00**
囪	0.93	0.78	**0.15**	0.90	0.70	**0.2**	0.90	0.85	**0.05**

[Table continued on next page.]

Table 7. [continued]

	Original			Strict			Semantics-favoring		
	Phon	Sem	**Dif**	Phon	Sem	**Dif**	Phon	Sem	**Dif**
亍	0.23	0.08	**0.15***	0.20	0.00	**0.20***	0.20	0.15	**0.05**
朮	0.23	0.00	**0.23***	0.20	0.00	**0.20***	0.20	0.00	**0.20***
冉	0.93	0.65	**0.28***	0.90	0.55	**0.35***	0.90	0.75	**0.15**
卞	0.88	0.58	**0.30***	0.85	0.55	**0.30***	0.85	0.60	**0.25**
弁	0.43	0.05	**0.38****	0.40	0.00	**0.40****	0.40	0.10	**0.30***

壬	0.93	0.45	**0.48*** **	0.85	0.30	**0.55*** **	0.85	0.60	**0.25**
兀	0.95	0.45	**0.50*** **	0.95	0.20	**0.75*** **	0.95	0.70	**0.25**
亙	0.90	0.38	**0.53*** **	0.85	0.20	**0.65*** **	0.85	0.55	**0.30**
丫	0.83	0.30	**0.53*** **	0.65	0.05	**0.60*** **	0.65	0.55	**0.10**

*0.01 < p < 0.05

**0.001 < p < 0.01

***p < 0.001

Unmarked: nonsignificant difference

According to the original coding, a majority of characters showing a significant difference between phonology and semantics had a higher accuracy for phonology, namely 朮, 弁, 亍, 卞, 亙, 兀, 壬, 丫, 冉; the only characters that had significantly higher accuracy for semantics were 廿 and 卅. With the strict coding, the pattern was similar, though the particular characters involved changed somewhat: 卮, 朮, 弁, 亍, 卞, 亙, 兀, 壬, 丫, 冉 all had significantly higher phonology scores, and only 廿 had a significantly higher semantics score. A quite different pattern arose with the semantics-favoring coding; as befits the name, almost all differences involved significantly higher semantics scores (卅, 孑, 卍, 叵, 仄, 廿), with only 弁 and 朮 having significantly higher phonology scores. Across the three codings, no character changed from one statistically significant pattern to the opposite, suggesting that all were tapping into the same basic phenomena, and three characters maintained their pattern regardless of the scoring system: 朮 and 弁 always had significantly higher phonology scores, and 廿 always had significantly higher semantics scores.

Note that most of the characters with higher phonology scores by at least one of the codings are used as phonological radicals, namely 卞, 亙, 兀, 壬, 冉, 卮; in addition, 丫 has a similar pronunciation as the zhuyin fuhao phonological transcription symbol ㄚ. The higher phonological accuracy for these is thus not particularly interesting, except that it suggests that participants may be able to guess pronunciations of unfamiliar simple characters by reference to complex characters that contain them, even if they are even less familiar. The appearance of 亍 on two of the lists of characters with higher phonology scores is more surprising, since its relationship with other characters is semantic, not phonological. Similarly, 朮 and 弁 show consistently higher phonology scores across all three coding systems, in spite of having no apparent relationships with other characters at all. However, the advantage for phonology with these characters doesn't imply that their phonology was really all that easy to access, as shown by their very low accuracy scores. It just so happened that the semantic scores for these items were even lower.

Gratifyingly, four of the five characters from our original set (廿, 卅, 卍, 孑) showed an advantage for semantics, though only for the first was this effect significant regardless of the coding system. The case of 孓 is rather more mysterious. Comparing its scores to those of 孑, with which it always appears (namely, in the word 孑孓), these characters had closely matched semantics scores across all three scoring systems, but the phonology scores for 孓 were noticeably higher. It is not clear to us why this should have been the case. There are two primary differences between 孑 and 孓 to keep in mind when considering possible explanations: the former can also appear in other lexical items while the latter cannot, and the former always appears in first position while the latter does not. One might expect that appearing in more than one lexical item should make a character's semantics easier to pinpoint, but that is not what happened here. Taft and Zhu (1995) found that accessing the pronunciation of a second-position binding character required first accessing that of the first-position character's, but this doesn't help explain our findings either, since this time it was the second-position character which had the more accurately recalled pronunciation. We are forced to leave this as an unexplained anomaly.

The remaining characters, which showed no special bias in favor of phonology or semantics, can be divided roughly into three categories depending on their overall scores. Of the three characters with overall high scores (both phonology and semantics scores above 0.5), one was a variant of a familiar character (万), and two were phonological radicals (囱 and 乍), though only one is associated with an obvious hint towards its meaning (namely, the appearance of 囱 in the synonymous compound 煙囱 /yan1cong1/ "chimney"); it's not clear what made the meaning of 乍 so familiar. None of the three characters with overall low scores (仝, 乜, 囟) were phonological radicals, though one of them (仝) is sometimes used as an alternate form for 同 /tong2/ "same". The apparent correlation between status as a phonological radical and overall accuracy on both phonology and semantics does not demonstrate a causal relationship, of course, since even assuming it holds up generally, it could be related to the inherent design of the Chinese orthographic system (i.e. there might be a sampling bias). Finally, the character 勺 was associated with scores in the middle range (around 0.5), even though it used as both a semantic and phonological radical.

Despite the inherent interest in the patterns found in the accuracy scores, in order to get a complete handle on what our participants were doing we also had to look at their self-assessments of confidence. These we turn to next.

4.3 Confidence Judgments

We began by performing by-participant and by-item ANOVAs on the mean confidence judgment scores in precisely the same way as we did for the accuracy scores. The patterns in confidence matched the patterns in accuracy in the most crucial way: the mean confidence score for pronunciation judgments was 3.42 (on a scale where 1 = least confident and 5 = most confident) while that for meaning judgments was 2.95 (because there were missing data points, the averages by participants and by items weren't identical; the values given here and elsewhere are the by-participant values). This was a significant difference both by participant ($F(1,18) = 8.85$, $p < 0.009$) and by item ($F(1,23)$

= 28.93, $p < 0.0001$). As with accuracy, both by-participant and by-item analyses failed to find a significant interaction in confidence scores between knowledge type and order (*F*s < 0.31, *p*s > 0.58). Thus, regardless of the order in which they had to make their judgments, participants were consistently more confident about the pronunciation than about the meaning.

Intriguingly, however, the order in which they made their judgments did have some effect on their overall confidence level: if they wrote down the meaning first, their mean confidence score across both types of knowledge was 3.34, while if they wrote down the pronunciation first, their mean confidence score across both types of knowledge was 3.02. However, this difference was only significant in the by-item analysis ($F(1,23) = 5.24$, $p < 0.032$), and not in the by-participant analysis ($F(1,18) = 0.75$, $p > 0.39$). It may be wisest not to read too much into this pattern, since it is conventional in psycholinguistics to accept effects as meaningful only if they are significant in both by-participant and by-item analyses, and this caution is especially relevant here, since our sample of items is quite atypical of the population Chinese characters as a whole. Even beyond the statistical technicalities, the difference in confidence that our participants may have felt was entirely illusory, since order had no effect on their accuracy scores, as we saw earlier. Nevertheless, the result is intriguing, since it hints that prior activation of character semantics may boost confidence about all aspects of a character, which is consistent with reading models that posit a direct semantic route. We will see later that a reflex of this pattern emerged from our signal detection analysis as well.

We also looked at differences in confidence judgments for all of the individual items. The results for these analyses are summarized in Table 8, with significance determined by paired two-tailed *t* tests.

Table 8. Differences in mean confidence scores (phonology - semantics); due to missing values, the number of participants providing data varied from item to item

	Number of participants	Pronunciation	Meaning	**Difference**
卅	20	3.30	3.95	**-0.65**
廿	19	3.32	3.65	**-0.33**
卍	18	3.00	3.00	**0.00**
万	20	3.05	3.00	**0.05**
亍	19	1.89	1.74	**0.16**
孓	20	3.95	3.70	**0.25**
仝	20	2.05	1.80	**0.25**
囟	19	1.95	1.63	**0.32**
孑	20	4.20	3.80	**0.40**
丫	20	3.80	3.35	**0.45***
勺	19	3.68	3.21	**0.47***
弁	19	2.70	2.21	**0.49**
壬	20	4.50	3.95	**0.55**
乜	19	2.40	1.84	**0.56***

[Table continued on next page.]
Table 8. [continued]

	Number of participants	Pronunciation	Meaning	**Difference**
卮	20	2.85	2.20	**0.65***
囪	20	3.65	3.00	**0.65**
兀	20	4.25	3.50	**0.75**
乍	20	4.25	3.50	**0.75*****
叵	19	2.95	2.15	**0.80****

仄	20	4.25	3.35	**0.90****
亙	20	4.00	3.05	**0.95****
卞	20	3.90	2.90	**1.00***
冉	18	4.55	3.44	**1.11*****
朮	19	3.35	2.00	**1.35*****

*$0.01 < p < 0.05$

**$0.001 < p < 0.01$

***$p < 0.001$

Unmarked: nonsignificant difference

It is clear from these values that our participants were consistently more confident about their knowledge of character pronunciations than about their knowledge of character meanings. The only characters that showed tendencies for greater confidence about semantics were 廿 and 卅, but even for these two, the differences were not statistically significant (for 廿, $t(18) = -0.55$, $p > 0.59$; for 卅, $t(19) = -1.51$, $p > 0.14$). As we saw earlier, both of these characters, along with a few others (孑, 卍, 叵, 仄), had relatively higher accuracy scores for semantics than for phonology, but this didn't lead to a statistically significant increase in confidence about this knowledge. In fact, the character 仄, which consistently showed higher accuracy scores for semantics than for phonology, nevertheless showed significantly higher confidence scores for phonology ($t(19) = 3.33$, $p < 0.004$). There thus seems to be a response bias in favor of phonology, and this bias might be obscuring any real differences in sensitivity towards semantics and phonology in characters.

In an attempt to tease apart sensitivity and bias in our task, we therefore decided to apply signal detection theory, as described in the next section.

4.4 Sensitivity and Bias

4.4.1 Background

Signal detection theory (see Macmillan and Creelman 2005 and Wickens 2002 for lucid introductions) was originally developed by psychophysicists studying low-level perception (e.g. detection of a faint light in a darkened room), but it is applicable to any situation where an observer is asked to detect something. We can think of the task that we gave to our participants in this way as well: they were asked to detect their internal representations of knowledge about a character's meaning and pronunciation. What we don't know, however, is whether their responses accurately reflect their actual ability to detect this knowledge (their sensitivity) or just their tendency to respond a particular way regardless of what they detect (their bias).

Signal detection theory provides tools for teasing these two factors apart. Its key insight is very simple. Namely, to determine whether experimental participants have really detected something just from observing their behavior, we must pay attention both to those occasions when they claim to detect something and they are right ("hits") and to those occasions when they claim to detect something but are in fact wrong ("false alarms"). We would not say that somebody is actually detecting a flashing light if she claims to see it in every experimental trial, even when there is no light at all. Rather, real sensitivity is only revealed if the hit rate is different from the false alarm rate: the greater the difference between these rates, the greater the sensitivity. Signal detection theory quantifies this insight in a sensitivity value called d' (its precise calculation involves first transforming the hit rate and false alarm rate in order to make them easier to interpret statistically). A value of $d' = 0$ implies that participants can't detect the signal at all, and nonzero values imply sensitivity. In the standard situation, sensitivity implies positive d', but if false alarms are more common than hits (as is likely to be the case in our experiment), negative d' values also indicate a sort of sensitivity, albeit to signals that the participants are misidentifying (similar to systematically claiming to see a light if and only if the light is actually off). The mathematics of *d'* means that perfect performance is associated with infinity, which can be avoided using various adjustments; in our analyses we follow Macmillan and Creelman (2005, pp. 8-9), and add 0.5 to all data cells before calculating hit and false alarm rates.

In contrast to sensitivity, a measure of bias should indicate the degree to which participants claim detection regardless of whether their response is a hit or a false alarm. One common way of quantifying this in signal detection theory is in a value that Macmillan and Creelman (2005) call *c* (functionally equivalent to the measure β used by other researchers; see Wickens 2002, pp. 26-31). This is based on the average of the hit rate and false alarm rate, transformed in such a way that *c* is zero if these two rates complement each other (i.e., if false alarm rates are equal to the miss rate, which is one minus the hit rate), and a nonzero value if they don't. A positive *c* means that the false alarm rate is lower than the miss rate, indicating bias *against* reporting a detection, while a negative *c* means the opposite, indicating a bias *for* reporting a detection. Sensitivity and bias are independent of each other, and have different properties. In particular, signal detection theory assumes that *d'* measures a participant's inherent ability to detect a signal, while bias may be shifted in various ways. For example, if you reward participants for increasing their hit rate, their *d'* value shouldn't change, while their *c* value should drop.

One way in which sensitivity and bias are they same, however, is that they are defined within individual participants; it makes no sense to calculate *d'* or *c* for individual items. Hence when looking for differences in sensitivity or bias for types of knowledge about characters, we can only perform by-participant analyses. This may seem like a serious failing for our purposes, given the important differences we uncovered across items in our experiment, but it is not in fact a fatal problem. Signal detection theory does not depend on having perfect balance across stimulus items. As long as some items do in fact have the signal to be detected and some do not, hit rates and false alarm rates can be calculated. Moreover, in our analyses, hit and false alarm rates were calculated separately for meanings and pronunciations, so it doesn't matter that accuracy and confidence for the latter tended to be higher; there were also many hits (i.e. correct responses) for meaning.

The application of signal detection theory to our results was of a somewhat more complex form than what we have just sketched, since rather than asking our participants simply to detect the presence of some type of knowledge about a character, we also asked them to give a scaled estimate of their level of confidence about their knowledge. Fortunately, there are also standardized procedures for analyzing rating experiments as well. The model that signal detection theory assumes for this is that each point on the rating scale represents a different degree of bias, not a different degree of sensitivity. Thus it is possible to estimate a single overall sensitivity value for each participant across all ratings; the specific estimate that we used was a sort of average d' that Macmillan and Creelman (2005, pp. 61-62) call d_a. A modification of the bias measure c is also necessary for rating experiments; we used c_a (see Macmillan and Creelman, 2005, p. 67). An experiment like ours, with a five-point scale, involves four c_a values, associated with biases to respond with a score no higher than 1, 2, 3, or 4, respectively; the model assumes that a score of 5 is given only when all of these biases are exceeded.

4.4.2 Analysis

In our use of signal detection theory, we counted a "hit" if a participant correctly identified the meaning or pronunciation of a character (according to some scoring system), and a "false alarm" if a participant wrote down an incorrect meaning or pronunciation (i.e. they claimed to detect knowledge which they didn't actually have). Two different scoring systems were used: the "strict" system and the "semantics-favoring" system, described above; the original three-point accuracy scale wasn't used since we needed to divide responses into two distinct categories. For each participant, a single sensitivity measure d_a was calculated, as well as four separate bias measures c_a for the four confidence judgment scores below 5.

We then put the d_a sensitivity values into by-participant two-way ANOVAs, with order as a between-group factor and knowledge type as a within-group factor. The mean d_a values for both phonology and semantics were negative (respectively, -1.03 and -0.95 for the strict scoring system, -1.03 and -1.03 for the semantics-favoring system), indicating that participants tended overall to misidentify their knowledge about characters (i.e. false alarms were more common than hits). There were no significant effects of order and no interactions between the two factors (Fs < 3.4, ps > 0.08). Moreover, regardless of scoring system, there was no significant effect of knowledge type (strict: $F(1,18) = 0.24$, $p > 0.63$; semantics-favoring: $F(1,18) < 0.0001$, $p > 0.99$). Thus in spite of their greater accuracy and confidence for phonological knowledge about characters, our participants did not in fact reveal any greater sensitivity to phonology compared with semantics.

Analyses of the c_a bias values required three-way ANOVAs, with order as a between-group factor and both knowledge type and rating score (labeled R1, R2, R3, R4) as within-group factors. Unsurprisingly, rating score was a highly significant factor, demonstrating that different rating scores were indeed associated with different degrees of bias (strict: $F(3,54) = 108.02$, $p < 0.0001$; semantics-favoring: $F(3,54) = 125.81$, $p < 0.0001$), with lower c_a values (indicating a greater bias for reporting a detection) associated with higher confidence ratings. While there was no main effect of order for either scoring system (Fs < 1.19, ps > 0.29), there was an interaction between rating score and order, though it only reached significance with the semantics-favoring rating (strict: $F(3,54) = 2.40$, $p > 0.07$; semantics-favoring: $F(3,54) = 3.02$, $p < 0.04$). Namely, as confidence ratings dropped (from the highest rating criterion down to the lowest), the effect of order on c_a increased, with c_a values consistently higher for participants who judged semantics first (except for the highest rating criterion R4, which showed no difference). Since higher c_a values imply a greater bias *against* reporting a detection, and since this effect was greater for lower ratings, it seems that judging semantics first made participants less likely to give low confidence scores overall, relative to what one would expect them to give based on their accuracy rates. This finding thus helps back up the observation noted earlier about the by-item analysis of the raw confidence ratings themselves, suggesting a special role for semantics in character processing in our task.

A more fundamental finding, however, was that using either scoring system, mean c_a values for phonology were higher than for semantics (respectively, 0.26 and 0.06 for the strict coding, 0.26 and -0.05 for the semantics-favoring coding), though again this pattern only reached statistical significance with the semantics-favoring coding (strict: $F(1,18) = 3.75$, $p > 0.06$; semantics-favoring: $F(1,18) = 8.68$, $p < 0.009$). Moreover, for both codings, the mean c_a value for phonology was significantly different from 0, while the mean c_a value for semantics was not significantly different from 0. Put together, these results indicate that participants had a phonological bias, but no semantic bias. Interestingly, the positive c_a values indicate that participants were biased *against* giving high confidence ratings for phonology. Thus, while the lack of a semantic bias indicates that the participants' low confidence scores for semantics accurately reflected their genuinely low levels of confidence, the positive phonological bias means that their confidence scores for phonology, already reported as higher than those for semantics, should have been even higher. That is, the participants were actually quite confident about their phonological knowledge (whether or not their confidence was justified by a real ability to detect this knowledge); they were, in a sense, just too modest to admit it.

5. CONCLUSIONS

This study has highlighted a number of facts relevant to understanding the roles of phonology and semantics in reading Chinese characters. Some of these we were forced to review merely through the exercise of preparing the experiment, as the inventory of Chinese characters was explored in ways that it often is not (at least in psycholinguistic studies). One of these facts is the high degree of interdependence across characters. While it is often noted that most characters are complex, less often discussed is the reverse situation: most simple characters are themselves radicals, allowing their meanings or pronunciations to be guessed by reference to complex characters. This two-way interdependence implies that Chinese orthography (even in the relatively more complex traditional form used in Taiwan) has evolved in response to memory load constraints; memorizing 5000 characters requires far less capacity than memorizing 5000 totally unrelated entities. A second basic fact about characters is that phonological radicals are more reliable indicators of pronunciation than semantic radicals are of meaning. As we have seen, this doesn't mean that this pattern necessarily arose because phonology is more fundamental than semantics in reading (it probably arose through the phonological biases of character inventors), but it is important to remember that the system itself has a decidedly phonological slant. Third, the continuing focus on characters in Chinese psycholinguistics obscures the fact that character recognition is just an early stage on the way to accessing higher level information like words, syntax, and propositions. Consideration of the word level, for example, reveals that access of character pronunciations independent of word context is inherently easier than access of character semantics, since characters almost always have the same pronunciation across words, while they very often vary in semantics.

Our experiment then probed readers' knowledge of the phonology and semantics of a collection of characters, which we chose because we expected either or both of these to be difficult to access. Superficially, the results seemed to strengthen the conclusion drawn from our descriptive analysis of the Chinese orthographic system: the experimental participants showed both higher accuracy and a greater degree of confidence about their knowledge of these characters' pronunciations than about their meanings. However, signal detection theory helped us to localize these effects in bias, rather than sensitivity, at least with respect to our particularly difficult collection of characters. The sensitivity analysis implied that readers actually knew no more about the pronunciations of these characters than their meanings, while the bias analysis implied that their confidence about phonology was higher than was justified by their sensitivity. These analyses undermine the assumption that higher scores for phonology necessarily mean that readers access phonological knowledge about characters more easily than semantic knowledge; it may seem that way to readers, but this is probably an illusion. We also observed signs that prior activation of character semantics may increase levels of confidence overall: the feeling of knowing a character, even if this feeling is illusory, is increased when semantic knowledge is probed first. These findings together imply that the role of phonology, even in leisurely pondering of isolated characters, may actually be less important than what is implied by the pro-phonology bias inherent in the structure of Chinese orthography.

Moreover, our experiment confirmed that there are indeed characters whose meanings are more accurately recalled than their pronunciations, such as 廿, 卅, 孑, 卍, 叵, 仄, though only the first of these showed this advantage consistently regardless of how we scored accuracy, and only the first two were associated with significantly higher degrees of confidence for semantics than for phonology. Nevertheless, the fact is that such characters do exist, so it is indeed possible for a character to be recognized for meaning without first activating the syllable associated with it.

Yet it must be admitted that, in spite of our initial skepticism, there are also characters that can be recognized for their pronunciation without activating any clear semantic representation: 朮 and 弁 were always associated with higher accuracy for phonology, even with the semantics-favoring accuracy scoring system. There are also reasons to be cautious when musing on the implications of characters with clearer meanings than pronunciations. As we discussed when introducing our materials, characters of this type have rather atypical characteristics when looked at within the total inventory of Chinese characters.

First and most obviously, they are simple and contain no phonological radical. Of course, this is why we selected them for the experiment, but it doesn't hurt to re-emphasize that our items did not form a representative sample. Second, some characters with significantly less familiar pronunciations are most typically used as graphical objects, rather than truly linguistic objects. This is clearly the case for 卍, and arguably also for 廿 and 卅 as well. That is, these characters have roughly the same status as symbols like "&" in Roman orthography. It seems likely that at least some English-speaking users of the symbol "&" may be fully aware of its meaning but are nevertheless unaware that its correct name is "ampersand". Yet it would be unfair to penalize them if they pronounced a bit of text like "you & I" as "you and I" rather than as "you ampersand I". In the same way, when 廿 and 卅 appear in text, it is entirely standard to pronounce them as /er4shi2/ and /san1shi2/, respectively.

It is rather surprising, in fact, given the traditional view of Chinese reading, how hard it is to find characters that are read primarily for meaning rather than pronunciation. In our study, the characters 孑, 叵, 仄 perhaps come closest to what we expected to find in abundance: characters associated with academic topics that one would usually only read about, but never hear being used in daily conversation. The polysyllabic and polymorphemic nature of modern Chinese, however, makes coming across such a character in a book or newspaper the exception rather than the rule.

Put together with previous research, then, our findings serve to reconfirm a nuanced position in the debates over the roles of semantics and phonology in Chinese reading: both are important, but in different ways.

NOTES

1. This research was supported by the Chiang Ching-Kuo Foundation (grant RG001-D-02). We greatly appreciate comments from an anonymous reviewer for helping us to improve the paper.

2. Since our experiment was conducted in Taiwan, all characters in this paper are given in their traditional form. While this may affect aspects of our specific observations, we expect the general principles to carry over to simplified characters as well.

REFERENCES

BALUCH, B., & BESNER, D. 1991. Visual word recognition: Evidence for strategic control of lexical and nonlexical routines in oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17* (4), 644-652.

BALUCH, B., & BESNER, D. 2001. Basic processes in reading: Semantics affects speeded naming of high-frequency words in an alphabetic script. *Canadian Journal of Experimental Psychology, 55* (1), 63-69.

BLAIR, I. V., URLAND, G. R., & MA, J. E. 2002. Using Internet search engines to estimate word frequency. *Behavior Research Methods, Instruments, & Computers, 34* (2), 286-290.

BYBEE, J. L., & MODER, C. L.. 1983. Morphological classes as natural categories. *Language,* 59, 251-270.

CHAO, Y.-R. 1968. *A grammar of spoken Chinese.* University of California Press.

CHEN, B., WANG L., & PENG, D. 2003. The time course of graphic, phonological and semantic information processing in Chinese character recognition (II). *Acta Psychologica Sinica, 35* (5), 576-581.

CHEN, B., & PENG, D. 2001. The time course of graphic, phonological and semantic information processing in Chinese character recognition (I). *Acta Psychologica Sinica, 33* (1), 1-6.

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.

CHEN, K.-J., HUANG, C.-R., CHANG, L.-P., & HSU, H.-L. 1996. SINICA CORPUS: design methodology for balanced corpora. *Language, Information and Computation, 11*, 167-176.

CHEN, Y.-P. 2002. Reading efficiency and reading strategies. In H. S. R. Kao, C.-K. Leong, & D.-G. Gao (Eds.), *Cognitive neuroscience studies of the Chinese language* (pp. 143-156). Aberdeen, Hong Kong: Hong Kong University Press.

CHOMSKY, N., & M. HALLE. 1968. *The sound patterns of English.* Cambridge, MA: MIT Press.

DEFRANCIS, J. 1984. *The Chinese language: Fact and fantasy*. Honolulu, HI: University of Hawaii Press.

DEFRANCIS, J. 1989. *Visible speech: The diverse oneness of writing systems*. Honolulu, HI: University of Hawaii Press.

FODOR, J. A. 1981. The present status of the innateness controversy. In J. A. Fodor (Ed.) *Representations*. Cambridge, MA: MIT Press.

General Chinese-English Dictionary. (1997). Hong Kong: Chung Hwa Book Co., Ltd.

HOOSAIN, R. 2002. Speed of getting at the phonology and meaning of Chinese words. In H. S. R. Kao, C.-K. Leong, & D.-G. Gao (Eds.), *Cognitive neuroscience studies of the Chinese language* (pp. 129-142). Aberdeen, Hong Kong: Hong Kong University Press.

HUNG, D. L., & TZENG, O. J. 1981. Orthographic variations and visual information processing. *Psychological Bulletin, 90*, 377-414.

KUO, W.J., YEH, T.C., DUANN, J. R., WU, Y. T., HO, L.T., HUNG, D. L., TZENG, O. T. L., & HSIEH, J. C. 2001. A left-lateralized network for reading Chinese words: an fMRI study. *NeuroReport, 12*(18), 3997-4001.

LI, H., LI, T.-K., & TSENG J.-F. (1997). *Guoyu cidian jianbianben bianji ziliao zicipin tongji baogao*. [Mandarin dictionary-based character and word frequency statistical report] Ministry of Education. Available at http://140.111.1.22/clc/dict/htm/pin/start.htm

LIANG, S.-Q. (Ed.). 1985. *Far East Chinese English Dictionary*. Taipei: Far East Book Co., Ltd.

MACMILLAN, N. A., & CREELMAN, C. D. 2005. *Detection theory: A user's guide*. Second Edition. Mahwah, NJ, USA: Lawrence Erlbaum Associates.

MARKMAN, E. M., & WACHTEL, G. F. 1988. Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*, 121-157.

MATTINGLY, I. G. 1992. Linguistic awareness and orthographic form. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 11-26). Amsterdam: Elesevier.

MYERS, J. To appear. Processing Chinese morphology: a survey of the literature. In G. Libben and G. Jarema (Eds.). *The representation and processing of compound words*. Oxford: Oxford University Press.

MYERS, J., DERWING, B., & LIBBEN, G. 2004. The effect of priming direction on reading Chinese compounds. *Mental Lexicon Working Papers, 1*, 69-86. University of Alberta, Canada.

PACKARD, J. L. 2000. *The morphology of Chinese: A linguistic and cognitive approach.* Cambridge, UK: Cambridge University Press.

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, & Cognition, 24*, 101-118.

PERFETTI, C. A., & TAN, L. H. 1999. The constituency model of Chinese word identification. In A. Inhoff, J. Wang, & H. C. Chen (Eds.), *Reading Chinese scripts: A cognitive analysis*. Mahweh, NJ: Erlbaum.

PERFETTI, C. A., & ZHANG, S. 1995. The universal word identification reflex. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 33, pp. 159-189). San Diego: Academic Press.

SHEN, H. H., & BEAR, D. R. 2000. Development of orthographic skills in Chinese children. *Reading and Writing, 13* (3-4), 197-236.

SU, I.-R. 2004. The effects of discourse processing with regard to syntactic and semantic cues: A competition model study. *Applied Psycholinguistics, 25* (4), 587-601.

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., & VAN GRAAN, F. 1998. Lack of phonological mediation in a semantic categorization task. *Journal of Memory and Language, 38*, 203-224.

TAFT, M., & ZHU, X. 1995. The representation of bound morphemes in the lexicon: A Chinese study. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 293-316). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

TAFT, M., & ZHU, X. 1997. Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23* (3), 761-775.

TAN, L. H., & PERFETTI, C. A. 1997. Visual Chinese character recognition: does phonological information mediate access to meaning? *Journal of Memory and Language*, *37*, 41-57.

TAN, L. H., HOOSAIN, R., & PENG, D.-L. 1995. Role of early presemantic phonological code in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 43-54.

TSAI, C.-H. 2000. *Mandarin syllable frequency counts for Chinese characters*. Retrieved March 12, 2004, from http://www.geocities.com/hao510/syllable/

TZENG, O. J. L., & HUNG, D. L. 1980. Reading in the non-alphabetic writing system: some experimental studies. In J. F. Kavanagh & R. L. Venezky (Eds.) *Orthography, reading, and dyslexia*. Baltimore, MD: University

Park Press.

TZENG, O. J. L., HUNG, D. L., & WANG, W. S-Y. 1977. Speech recoding in reading Chinese characters. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 621-630.

WICKENS, T. D. 2002. *Elementary signal detection theory*. Oxford: Oxford University Press.

ZHANG, G., & SIMON, H. A. 1985. STM capacity for Chinese words and idioms: Chunking and the acoustical loop hypotheses. *Memory and Cognition, 13*, 193-201.

ZHOU, X., & MARSLEN-WILSON, W. 1994. Words, morphemes, and syllables in the Chinese Mental lexicon. *Language and Cognitive Processes, 9* (3), 393-422.

ZHOU, X. & MARSLEN-WILSON, W. 1999. Phonology, orthography, and lexical semantic activation in reading Chinese. *Journal of Memory and Language, 41*, 579-606.

ZHOU, X., & MARSLEN-WILSON, W. 2002. Semantic processing of phonetic radicals in reading Chinese characters. *Acta Psychologica Sinica, 34* (1), 1-9.

ZHOU, X., SHU, H., BI, Y., and SHI, D. 1999. Is there phonologically mediated access to lexical semantics in reading Chinese? In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 135-171). Mahwah, New Jersey: Lawrence Erlbaum Associates.

排除意義或發音媒介的漢字辨識

James Myers*, Marcus Taft**, 周培瑛*

*國立中正大學　****University of New South Wales***

在中文的心理語言學研究上，語意和語音在閱讀中的相對角色一直是長久以來的爭議。有些研究者認為閱讀漢字時需要音韻表徵的作用；而有些仍持傳統看法，認為讀者可以直接提取語意，不需要音韻作為媒介。本實驗從一個

創新的角度來探討這個議題。我們請受試者報告他們對不熟悉意義或發音的獨體漢字的瞭解。依照「音韻優先」的看法，知道意義而不知道發音是不可能的；依照「語意優先」的看法，則知道發音而不知道意義是不可能的。我們的實驗結果顯示這兩種情況都可能存在，但有質與量的不同。知道發音而不知道意義是比較普遍的情形，通常發生在有同樣聲符的漢字；而知道意義卻不知道發音的情形通常發生在當受試者認為該漢字還有其它發音，或此漢字實際上是代表某個字的符號，並不是真的語言。此外，在訊息偵測分析中，我們發現語意和發音在敏感度上並無不同。然而，受試者卻對音韻判斷有較強的信心。在中文閱讀研究的爭議中，我們的實驗結果再度肯定這個更精細與周全的立場：音韻和語意兩者都在閱讀上扮演著關鍵的角色。

關鍵詞
漢字，書寫系統，讀取，音韻，語義，心理語言學