American Psychological Association

Journal of Experimental Psychology: Human Perception and Performance

ISSN: 0096-1523

2019, Vol. 45, No. 11, 1513–1521
http://dx.doi.org/10.1037/xhp0000692

On Recognizing Japanese Katakana Words: Explaining the Reduced Priming With Hiragana and Mixed-Kana Identity Primes

Sachiko Kinoshita
Macquarie University

Rinus G. Verdonschot
Hiroshima University

The Japanese kana syllabary has 2 allographic forms, hiragana and katakana. As with other allographic variants like the uppercase and lowercase letters of the Roman alphabet, they show robust form-independent priming effects in the allograph match task (e.g., Kinoshita, Schubert, & Verdonschot, 2019), suggesting that they share abstract character-level representations. In direct contradiction, Perea, Nakayama, and Lupker (2017) argued that hiragana and katakana do not share character-level representations, based on their finding of reduced priming with identity prime containing a mix of hiragana and katakana (the mixed-kana prime) relative to the all-katakana identity prime in a lexical-decision task with loanword targets written in katakana. Here we sought to reconcile these seemingly contradictory claims, using mixed-kana, hiragana, and katakana primes in lexical decision. The mixed-kana prime and hiragana prime produced priming effects that are indistinguishable, and both were reduced in size relative to the priming effect produced by the katakana identity prime. Furthermore, this pattern was unchanged when the target was presented in hiragana. The findings are interpreted in terms of the assumption that the katakana format is specified in the orthographic representation of loanwords in Japanese readers. Implications of the account for the universality across writing systems is discussed.

Public Significance Statement

Japanese kana syllabary has two allographic variants, hiragana and katakana. They have been shown to behave like the uppercase and lowercase variants of the Roman alphabet in some experiments but not in others. The present paper provides a coherent account of these seemingly contradictory data, borrowing from the orthographic cue hypothesis, the idea that orthographic format of some types of words serves as linguistic cues. Researchers should be mindful of the letter-word equivalence assumption, the danger in using word stimuli to draw conclusions about letter-level representations.

Keywords: masked priming, allograph, Japanese hiragana and katakana, orthographic cue hypothesis, lexical decision

A written word is comprised of a sequence of letters, with the identity of each letter determined by its visual form. Some letters are visually similar and are distinguished only by a very small difference in visual features, for example: O and C; b and p. At the same time, there are different visual forms that correspond to the same letter identity for example, a and A, r and R. Form variants corresponding to the same letter identity are called allographs.[1] The Roman alphabet is not the only writing system with allographic variation: Greek and Cyrillic alphabets also have upper- and lowercase variants; the Arabic alphabet has position-specific allographs (in which a letter may take up to four different forms depending on the position of the letter within a word); the Japanese kana syllabary has two allographic forms, hiragana and katakana.

In becoming literate with writing systems with allography, the reader learns to map the allographic variants (e.g., a and A) onto a common abstract letter identity (ALI). Support for the ALI comes from many sources, including neuropsychological data

This article was published Online First August 22, 2019.

Sachiko Kinoshita, Department of Psychology and Macquarie University Centre for Reading, Macquarie University; Rinus G. Verdonschot, Department of Oral and Maxillofacial Radiology, Institute of Biomedical and Health Sciences, Hiroshima University.

Sachiko Kinoshita's research is supported by an ARC Discovery Project Grant (DP140101199). Rinus G. Verdonschot's research is sponsored by Grant-In-Aid (C) 17K02748 from the Japanese Society for the promotion of Science. The raw data from this study can be found on the Open Science Framework, at the following URL: https://osf.io/3cn7s/.

Correspondence concerning this article should be addressed to Sachiko Kinoshita, Department of Psychology and Macquarie University Centre for Reading, Macquarie University, Sydney, New South Wales 2109, Australia. E-mail: sachiko.kinoshita@mq.edu.au

[1] A different perspective (see e.g., the third entry in Wikipedia, https://en.wikipedia.org/wiki/Allography) defines an allograph as "a smaller fragment of writing, that is a letter or a group of letters, which represents a particular sound. In the words *cat* and *king, the letters c and k are both allographs of the same sound*" (emphasis added). This phonology-centered perspective on allography is not one taken in contemporary experimental and neuropsychological studies of orthographic processing; cat and kat are homophonic but c and k are not considered to be allographs.

from impaired readers (e.g., Coltheart, 1981; Schubert & McCloskey, 2013) and neural data (e.g., Dehaene et al., 2004) as well as behavioral data with skilled adult readers. Many of the behavioral studies (e.g., Bigsby, 1988) have made use of the allograph match task (also called the cross-case letter match or nominal identity match), in which participants are asked whether a pair of letters (e.g., a and A) are the same or different letter, irrespective of letter case. More recently, Norris and Kinoshita (2008) combined the masked priming procedure with the allograph match task and showed that robust priming effect is observed that is independent of visual similarity: that is, visually dissimilar allograph pairs (e.g., a and A, r and R) produce equal-sized priming as visually similar allograph pairs (e.g., c and C, x and X; see also Kinoshita & Kaplan, 2008 for a discussion of why previous studies using the alphabet decision task have failed to find the form-independent priming effect). This finding has been extended to Arabic allographs (Carreiras, Perea, & Abu Mallouh, 2012), and these authors went as far as to suggest that priming of abstract letter identities may be universal.

In acquiring literacy, the reader learns not only the allographic variants that map onto the same abstract letter identity but also the linguistic conventions associated with the use of distinct allographic variants. For example, in English and other European languages written in the Roman alphabet, the initial letter is capitalized when writing proper names like personal names and geographical place names (e.g., Elizabeth, America); abbreviations (e.g., CIA, IBM) are typically written all in uppercase; and in German, nouns are initial capitalized (e.g., das Auto, the car). In Arabic, a literate reader would know which allographic variant is used in which specific position (initial, middle or final) within a word. These conventions regarding the use of allographic variants almost always concern a sequence of letters or a word, not a single letter in isolation. Earlier, Bowers, Vigliocco, and Haan (1998) cautioned researchers working on visual word recognition against what they called the letter-word equivalence assumption. Specifically, they noted that "researchers typically investigate word (or pseudoword) identification, and the results are assumed to have implications for letters" (p. 1705), an assumption that may not be warranted. The pitfalls of letter-word equivalence assumption are particularly acute in investigations of foreign language/writing system because it is easy to misinterpret/misattribute the source of an effect here. The present paper focuses on one such instance, concerning the role of allography in the Japanese kana syllabary, and begins with a description of the kana writing system.

Japanese Kana System

Japanese text is written using two distinct writing systems, logographic kanji and syllabic kana. Japanese kanji were adopted from Chinese hanzi, in which each character is a morpheme. Kanji is used to write content words such as nouns (e.g., 星,/ho.shi/, star) and stems of adjectives and verbs (e.g., 美しい/*u.tsu.ku*.shi.i/beautiful or 止まる/*to*.ma.ru/to stop; in both examples the first character is kanji). There are 2,136 official kanji (jôyô kanji) prescribed by the Japanese Ministry of Education, which are taught during the 9 years of compulsory education (see Tamaoka, Makioka, Sanders, & Verdonschot, 2017), although there are many more kanji in existence (kanji dictionaries may contain well more than 12,000 entries, cf. Chikamatsu, Yokoyama, Nozaki, Long, & Fukuda, 2000).

Kana is a syllabary and there are only 46 basic kana characters. Kana is learned earlier than kanji: Japanese children are formally taught kana in the first year of primary school, and most children master the whole set of kana by the end of first year; and in practice, many children already know (some) kana prior to starting school. Each kana character denotes a mora, a syllable-like phonological unit comprised of either a single vowel or a consonant-vowel combination (with the exception of a nasal consonant mora, /n/, which cannot be word-initial). Unlike the alphabetic system, there is no distinction between letter names (e.g., "tee" for t/T) and letter sounds (the phoneme /t/); the mora serves as both. For each mora, there is a hiragana form and katakana form: For example, け and ケ are the hiragana and katakana characters, respectively, for the mora /ke/. Hiragana and katakana are allographs analogous to the upper- and lowercase forms in the Roman alphabet, that is, two parallel forms that have different visual forms but correspond to the same letter identity and have the same letter name. Consistent with this, just as with the Roman alphabet in the cross-case letter match, Japanese readers have no difficulty in matching hiragana and katakana forms (i.e., deciding that け and ケ are the same), and—paralleling the findings with the Roman alphabet—Kinoshita et al. (2019) showed that Japanese readers show robust cross-kana identity priming effect for visual dissimilar hiragana/katakana pairs.

One feature of the Japanese kana system is that hiragana and katakana are used to write different types of words. Hiragana is used mainly to write inflections of adjectives and verbs (e.g., 美しい/*u.tsu.ku*.shi.i/beautiful, 止まる/*to*.ma.ru/to stop; in both examples the last two characters are hiragana) and case markers; additionally, content words of Sino-Japanese origin for which kanji are rare or unfamiliar (e.g., そろばん/so.ro.ba.n/abacus). Katakana, on the other hand, is used almost exclusively to write loanwords, usually nouns of western origin (e.g., ラジオ/ra.ji.o/ radio). Whereas in principle any spoken word (or nonword) can be written in hiragana or katakana, loanwords are almost always written in katakana and rarely appear in hiragana.

The fact that hiragana and katakana are used to write different types of words is a main factor[2] that motivated Perea et al. (2017) to conduct their experiment that led them to conclude that "it does

[2] Perea et al. (2017) described "another salient feature" (p. 1141) of Japanese katakana and hiragana as being different from the uppercase-lowercase distinction in the Roman alphabet. They wrote that "no words are ever written in a combination of two scripts" (p.1141), and hence, "when one encounters contiguous characters in hiragana and katakana, (it) implies that readers are encountering a word boundary" (p.1142). This is incorrect. As noted, hiragana (but not katakana) is used to write inflections, and thus, an inflected loanword adjective is written as a combination of katakana (the stem) and hiragana (inflection) e.g., エレガントな (/e.re.ga.n.to.na/, elegant); ラフな (/ra.fu.na/, rough) (in both examples the last character な(/na/) is hiragana and the rest are written in katakana). Another example is the honorific suffix さん (/sa.n/), which is always written in hiragana (note that unlike an honorific in English (e.g., Mr., Mrs.); this cannot stand alone (i.e., it is a bound morpheme). This suffix can be used in combination with a foreign name written in katakana, e.g., ラプカーさん (/ra.pu.ka.a.sa.n/, Mr. Lupker). Thus, contiguous hiragana and katakana characters may correspond to a morpheme boundary but not necessarily a word boundary, contrary to Perea et al.'s (2017) claim, and hence, what they described as the "salient feature of Japanese" (p.1141) does not provide a rationale for expecting the alternating kana manipulation to behave differently from alternating case manipulation in the Roman alphabet.

seem that the characters in the two scripts do not share abstract character-level representations" (p.1145). It is this claim and the data used to argue for the claim that form the starting point of the present research. We point out two issues with Perea et al.'s (2017) claim.

The first relates to the logic to motivate the claim. Perea et al. (2017) noted the fact that hiragana and katakana are used to write different types of words to suggest that "logically, a reasonable argument could be made that katakana and hiragana characters will not become mapped onto the same character-level representation" (p.1141). As support for this possibility, Perea et al. referred to a suggestion made by Polk et al. (2009) that abstract letter representations may develop as a result of readers associating a set of visually very different letters occurring in the same word context: For example, the allographs a and A occur in the company of letters C and P in the word context CAP and cap. However, this line of argument requires that the mapping of allographic variants to a shared ALI is learned only in the context of words, a possibility that Polk et al. (2009) explicitly raised and rejected. Specifically, Polk et al. wrote: "Although our results suggest that visual context plays an important role in the acquisition of ALIs, they do not prove that other kinds of nonvisual information (e.g., the sounds of the letters, the meanings of the words) play no role. *Young children often know quite a lot about letters in isolation, including their names, before they have significant experience reading words, and it is certainly plausible that this knowledge plays a role in the acquisition of ALIs*." (p. 89, emphasis ours). Indeed, literacy instruction usually starts with the teaching of the letter/character inventory as isolated single letters/characters in a specific order (i.e., learning your abcs), and this is also the case with the kana syllabary (i.e., the equivalent of alphabetical order is called the a-i-u-e-o order). The mapping of hiragana and katakana onto the same character representation could be learned during the learning of kana inventory. Consistent with this, Schubert, Gawthrop, and Kinoshita (2018) found that adult L2 learners of Japanese (who had only a few semesters of Japanese language instruction) showed robust cross-kana priming in an allograph match task and concluded that the ALIs are acquired relatively early. In sum, there is little support—either logical or empirical—to motivate the suggestion that the distinct use of hiragana and katakana to write different types of words prevents the development of shared abstract kana identities at the character level.

A second issue with Perea et al.'s (2017) claim that hiragana and katakana characters do not share abstract character-level representations concerns the interpretation of their data. The data came from a lexical decision experiment using words that are normally written in katakana (e.g., レストラン, roman transcription/re.su-.to.ra.n/, meaning restaurant) as targets. Perea et al. compared the katakana identity prime (the same-script prime) and mixed-kana identity prime in which the second and fourth characters were written in hiragana (e.g., レすトらン) and found the priming effect was reduced for the latter. This result contrasts with the finding of lexical decision studies with languages written in the Roman alphabet that for targets written in uppercase letters masked identity priming effects of similar size are found with the lowercase, uppercase, and mixed-case primes (e.g., beard-BEARD; BEARD-BEARD; and BeArD-BEARD; e.g., Forster, 1998; Jacobs, Grainger, & Ferrand, 1995; Perea, Vergara-Martínez, & Gomez, 2015), which is widely interpreted in terms of the involvement of ALIs. Based on this contrast, Perea et al. (2017) concluded that in Japanese, hiragana and katakana do not share character-level representations. However, Perea et al.'s (2017) conclusion is contradicted by the finding of robust cross-kana identity priming effects at the single character level: As mentioned earlier, using the masked priming allograph match task, Kinoshita et al. (2019) showed a robust identity priming effect with kana primes written in the opposite kana format (e.g., い – イ or イ – い) and also that this priming effect could not be explained in terms of shared phonology. The finding of robust priming effects for allographs (presented as single characters) that differ in form parallels the findings with the Roman alphabet (e.g., Kinoshita et al., 2008; Norris et al., 2008) and Arabic (e.g., Carreiras et al., 2012) described earlier, consistent with the view that priming of abstract letter representations is language universal (Carreiras et al., 2012). The issue here is that Perea et al.'s (2017) data were based on word stimuli, but their conclusion concerned character-level representations—that is, their conclusion that katakana and hiragana characters do not share abstract character-level representations may represent an unwarranted application of letter-word equivalence assumption cautioned against by Bowers et al. (1998).

If, like the lowercase and uppercase variant of the Roman alphabet, hiragana and katakana share ALIs, how then can the reduced priming with mixed-kana identity prime reported by Perea et al. (2017) be explained? Kinoshita et al. (2019) suggested that this may have been due to the unfamiliar orthographic format of a mixed-kana prime. A Japanese loanword (which is almost always written in katakana) written in a mix of hiragana and katakana is orthographically unusual, like an abbreviation written in lowercase (e.g., dna, erp) or proper name written all in lowercase (america). Peressotti, Cubelli, and Job (2003) have found that lexical decision to proper name targets was indeed slower when they were presented in lowercase than when they were initial capitalized (e.g., america > America) and took the result to suggest that for proper names, information about the case of the initial letter is specified in the orthographic representation stored in the reader's long-term knowledge, a possibility they referred to as the orthographic cue hypothesis. As we noted earlier, a literate reader knows not only the allographic variants that map onto the same abstract letter identity (e.g., a = A; け = ケ) but also the linguistic conventions associated with the use of distinct allographic variants. Our suggestion therefore is that similarly to Peressotti et al.'s (2003) suggestion for proper names in the readers of the Roman alphabet, for Japanese readers, the katakana format is specified in their stored orthographic representation of loanwords. Accordingly, when an identity prime is presented in an orthographically unfamiliar format (e.g., a mix of hiragana and katakana), it does not match the orthographic description of the stored representation, hampering the process of finding a representation in the lexicon that matches the written input (i.e., orthographically driven lexical access) for these words.

In putting forward the above account, Kinoshita et al. (2019) noted that there is one extant finding that is inconsistent with their account. In an earlier study, very similar to Perea et al. (2017) and also using katakana words as targets (e.g., ピーマン,/pi.i.ma.n/, pepper) in lexical decision, Pylkkänen and Okano (2010) reported that the orthographic format of a masked prime did not affect the size of priming: Specifically, the identity prime in which the target was transcribed in hiragana (e.g., ぴーまん) produced an equal-sized priming effect as the katakana identity prime (e.g., ピーマン). A hiragana-transcribed loanword is also orthographically unfamiliar, so according to our orthographic cue account, hiragana identity primes should also have produced reduced priming, just like the mixed-kana identity prime.

In discussing this inconsistency between the effects of mixed-kana and hiragana identity primes, Kinoshita et al. (2019) speculated that the null difference in the size of priming produced by katakana and hiragana identity primes observed by Pylkkänen et al. (2010) may have been a type II error because of a lack of power because their experiment involved only eight participants. Also, Pylkkänen et al. (2010) and Perea et al. (2017) used different stimuli as well as different participants. Given this, the best approach to resolve the empirical inconsistency would be to replicate the two studies within a single experiment. This then was what we did in our Experiment 1.

Experiment 1

Method

Participants. Twenty-four students (10 male, mean age 22.74 years, $SD = 1.99$) from Hiroshima University participated in the experiment in return for 500 yen (US$4.40). All were native speakers and readers of Japanese and had normal or corrected-to-normal vision.

Design. The experiment was a masked priming lexical-decision task. The targets were preceded by one of four types of primes: (a) katakana identity, (b) hiragana identity, (c) mixed kana identity, and (d) control. The dependent variables were response latency and error rate.

Materials. The critical stimuli were 120 five-character words typically written in katakana, used by Perea et al. (2017). (We are grateful to them for sharing the stimuli.) Each target was preceded by one of four prime types: (a) katakana identity prime (e.g., レストラン - レストラン), (b) hiragana identity prime (e.g., れすとらん - レストラン), (c) mixed kana identity prime in which the first, third, and the fifth character was in hiragana (e.g., れスとらん - レストラン), and (d) control, which was another katakana word that shared no characters in the same position as the target (e.g., イヤリング - レストラン).

The 120 words were divided into four sets matched on mean frequency, for the purpose of counterbalancing the assignment of items to one of the four prime type conditions. Latin square was used to assign the four sets to the four prime conditions so that each word appeared in a list once, and across the four lists, it appeared in all four prime type conditions. Participants were assigned to one of four lists, in the order of arrival.

In addition to the 120 critical words, there were 120 five-character nonwords written in katakana. They were also identical to the nonword stimuli used by Perea et al. (2017). The same four types of primes used with the word targets were generated for the nonword targets, and four counterbalanced lists were generated for the nonword targets. There were also 10 words and 10 nonwords generated similarly and used for practice. The practice items were not included in the analysis.

Apparatus and procedure. Participants were tested individually or in pairs, seated approximately 60 cm in front of a computer monitor, upon which stimuli were presented. Each participant completed 240 test trials, in two blocks with a self-paced break between the two blocks. Within each block, a different random order of trials was generated for each participant. A practice block of 20 trials preceded the test block.

The task was lexical decision. Participants were instructed at the outset of the experiment that on each trial they would be presented with a character string written in katakana, and their task was to decide whether it was a word, as fast and accurately as possible. They were instructed to press the right shift key for Word and the left shift key for Nonword response.

Stimulus presentation and data collection were achieved through the use of the DMDX display system (Forster et al., 2003). Stimulus display was synchronized to the screen refresh rate (16.7 ms).

Each trial started with the presentation of a forward mask consisting of five # signs, for 500 ms. It was then replaced by the prime for 50 ms, followed immediately by a target. The target was presented in the same font (Hiragino Maru Gothic) as the prime and was magnified 1.2 times to avoid physical overlap with the identity prime. Targets were presented for a maximum of 2,000 ms, or until the participant's response. Participants were given an error feedback (a red X) when an error was made.

Results

Correct reaction times (RTs) and error rates for the word targets were analyzed using linear mixed-effects (LME) modeling with subjects (24) and items (120) as crossed random factors (Baayen, 2008), using the packages lme4 (version 1.1-17, Bates, Maechler, Bolker, & Walker, 2018), and lmerTest (version 3.0-1, Kuznetsova, Brockhoff, & Christensen, 2018) implemented in R (version 3.5.1, 2018-07-02, R Core Team, 2018). The fixed effect factors frequency (log-transformed) and Prime type, referenced to the mixed kana condition. In the analysis of RTs, errors trials (117 trials, 4.1%) were excluded, and the RTs were log-transformed to meet the distributional assumptions of LME. Upon the inspection of Q-Q plot, two data points with RT slower than 1,300 ms were excluded as outliers, resulting in 2,761 data points for analysis. We tested linear mixed-effects models with maximal random effects structures and simplified the random effects structure if the model did not converge or the model fit was not improved by model complexity. The raw data and the analysis file from this study can be found on the Open Science Framework, at the following URL: https://osf.io/3cn7s.

In addition, to quantify the amount of evidence for the critical effects, we calculated Bayes factors using the BayesFactor package (version 0.9.12-4.1, Morey & Rouder, 2018). A Bayes factor is an odds ratio, with 1 indicating equal evidence for the two alternative hypotheses, and generally odds of 3 or greater is considered to provide some evidence, greater than 10 to be strong evidence, and odds greater than 30 to be very strong evidence (Dienes, 2014; Jeffreys, 1961).

RT. The mean correct RT and error rates are shown in Table 1. The final statistical model we report here is: logRT ~ logfreq + Primetype + (1 | word) + (1 | subj), with the Primetype referenced to the mixed kana identity prime condition. The effect of frequency was significant, $t = -7.701$, $B = -.032$, $SE = .004$, $p < .001$. The control condition was significantly slower than the mixed kana identity prime condition, $t = 9.626$, $B = .084$, $SE = .009$, $p <$

Table 1
Mean Lexical Decision Latencies (RT, in Ms) and Percent Error Rates in Experiments 1 and 2

Target kana format		Experiment 1 Katakana (e.g., レストラン)		Experiment 2 Hiragana (e.g., れすとらん)	
Prime type	Example	RT (%E)	Priming	RT (%E)	Priming
Word targets					
Katakana identity	レストラン	530 (4.6)	72 (1.2)	627 (4.3)	91 (7.4)
Hiragana identity	れすとらん	544 (2.2)	58 (3.6)	651 (9.0)	67 (2.7)
Mixed kana identity	れスとラん	550 (3.6)	52 (2.2)	656 (7.9)	62 (3.8)
Control	イヤリング	602 (5.8)		718 (11.7)	
Nonword targets					
Katakana identity	ラスノモス	604 (1.7)	10 (−.6)	752 (2.5)	−1 (−.1)
Hiragana identity	らすのもす	626 (1.4)	−12 (−.3)	737 (2.5)	14 (−.1)
Mixed kana identity	らスのモす	614 (1.9)	0 (−.8)	741 (1.9)	10 (.5)
Control	ハレクカノ	614 (1.1)		751 (2.4)	

Note. %E = percent error rates.

.001, indicating that the mixed kana prime produced a significant priming effect. However, the katakana identity prime condition was significantly faster than the mixed kana identity prime condition, $t = -4.647$, $B = -.042$, $SE = .009$, $p < .001$, indicating that the priming effect produced by the mixed kana identity prime is smaller than that produced by the katakana identity prime. The pattern of priming produced by the mixed kana identity prime replicated the pattern reported by Perea et al. (2017). The Bayes factor for the difference between the katakana identity and mixed kana identity prime was 26,904 against the null model, indicating very strong support for the reduction in the size of priming effect produced by the latter. Of interest, the hiragana identity prime condition did not differ significantly from the mixed kana identity condition, $t = -1.518$, $B = -.013$, $SE = .008$, $p = .13$. The Bayes factor for this comparison was about 5 in favor of the null model. Referenced to the hiragana identity prime condition, the control condition was significantly slower, $t = 8.559$, $B = .099$, $SE = .012$, $p < .001$, and the katakana identity prime condition was significantly faster, $t = -3.608$, $B = -.030$, $SE = .008$, $p < .001$. The Bayes factor for the comparison between the katakana and hiragana identity prime condition was 67 against the null model, indicating strong evidence for the reduction in priming in the hiragana identity prime condition. Thus, the hiragana identity prime produced the pattern of priming that was indistinguishable from that produced by the mixed kana identity prime, that is, statistically significant but reduced in size relative to the katakana identity prime.

Error rate. Error rates were analyzed with generalized linear mixed-effects model with subjects and stimuli as crossed random factors, using the logit function appropriate for categorical variables (Jaeger, 2008). The statistical model we report here is: error ~ logfreq + Primetype + (1 | word) + (1| subj), with the Primetype referenced to the mixed kana identity prime condition. Frequency significantly reduced errors, Z = −4.570, $B = -.632$, $SE = .138$, $p < .001$. Relative to the mixed kana identity prime, only the control condition differed significantly, Z = 2.163, $B = .603$, $SE = .279$, $p < .05$. We then ran the same model referenced to the hiragana identity prime condition. Here the control condition was significantly more error prone, Z = 3.654, $B = 1.182$, $SE = .323$, $p < .001$, as was the katakana identity prime condition, Z = 2.614, $B = .871$, $SE = .333$, $p < .01$.

Discussion

Experiment 1 used the katakana target words used by Perea et al. (2017) and compared the size of priming by the katakana, hiragana, and mixed-kana identity primes within a single lexical decision experiment. The results were clear. The experiment replicated the reduced priming effect produced by the mixed kana identity prime relative to the katakana identity prime reported by Perea et al. (2017, Experiment 1) and, importantly, showed that the hiragana identity prime also produced a reduced priming effect, which was indistinguishable in size from the effect produced by the mixed-kana identity prime. The Bayes factor analysis provided strong evidence for the reduction in priming for both the mixed-kana prime and hiragana identity prime relative to the katakana prime and moderate support for the null difference between these two orthographically unfamiliar prime conditions. Taken together, the present results are entirely consistent with the interpretation that the identity priming effect for katakana words is reduced when the prime is written in an orthographically unfamiliar format.

The reduced priming by hiragana targets observed here is of course at odds with Pylkkänen et al.'s (2010) finding of equivalent priming effects with hiragana and katakana identity primes. The question is why Pylkkänen et al. (2010) did not observe a reduced priming effect with their hiragana identity primes. We have speculated earlier that their finding may have been a type II error, and here we consider several specific reasons. One is participant characteristics. Pylkkänen et al.'s (2010) participants were all residents of New York at the time of testing, and their reading of Japanese, and more specifically their sensitivity to the familiarity of orthographic format, may not have been on par with the Japanese university student participants tested here and in Perea et al.'s (2017) experiment. Consistent with this, their lexical decision latencies were unusually slow: The fastest (katakana identity prime) condition was much greater than 700 ms, compared with the 530 ms observed here and 520 ms in Perea et al. Another factor concerns their stimuli: Several reasons may be suggested as to why their katakana words did not appear as orthographically unfamiliar when transcribed in hiragana. First, a few of their katakana words are not a loanword of Western origin and could plausibly be written in hiragana (e.g., ごきぶりcockroach, おおかみ wolf, こんろ stove). Second, several of Pylkkänen et al.'s (2010) words contained characters that have the same form in katakana and hiragana, weakening the effect of format change (akin to using letters that are visually similar in uppercase and lowercase e.g., six-SIX). One such character is the へ (he) character (e.g., モヘア/mo.he.a/, mohair, メルヘン/me.ru.he.n/, fairy tale in German). Another is the dash character (e.g., ー in ピーマ ン, pepper). Although this character occurs only in the katakana format,[3] Pylkkänen et al. kept the dash character in the hiragana-transcription (e.g., ピーマ ン -> ぴーまん). In contrast, Perea et al. (2017) avoided words containing the へcharacter and the dash character. Finally, the target words used by Pylkkänen et al. (2010) were short, ranging from two to four kana characters (in contrast, Perea et al.'s stimuli were all five characters long), which would necessarily limit the scope of orthographic format manipulation implemented by transcribing the characters into hiragana. In sum, both the participant and stimulus characteristics of Pylkkänen et al.'s (2010, Experiment 1) study may have mitigated against finding reduced priming with hiragana primes relative to katakana primes.

Experiment 2—Targets in Hiragana

Experiment 1 showed that the identity priming effect for katakana words is reduced both when the prime was written in hiragana and in a mix of katakana and hiragana, relative to the katakana identity prime. We interpreted this pattern in terms of the assumption that the katakana format is specified in the orthographic representation of loanwords in the mental lexicon of Japanese readers: Relative to a katakana identity prime, an identity prime in an unfamiliar orthographic format (hiragana or a mix of hiragana and katakana) would therefore not match this orthographic description, thus reducing the facilitation in accessing this representation (i.e., priming). Another interpretation of the data, however, is that the katakana identity prime produced greater priming because it shared more characters with the target. The result of Perea et al.'s (2017) Experiment 2 seems consistent with this possibility. In this experiment the priming effect produced by the mixed kana identity prime (e.g., レ すトらン) was the same size as that produced by partial primes that contained only the katakana characters in the mixed-kana primes and had the hiragana characters replaced with asterisks (e.g., レ *ト*ン).

To test between these possibilities, we conducted Experiment 2. Here the prime conditions remained the same as Experiment 1, but the target was presented in hiragana (e.g., れすとらん). This means that the hiragana identity prime now contained the same characters as the target, and the katakana identity prime shared none. Hence, if the number of shared characters is what determines the size of priming, then here the hiragana prime should produce a larger priming effect than the katakana identity prime. In contrast, according to our orthographic cue account, the pattern of priming should remain the same, with the katakana identity prime producing priming effect of larger size.

Method

Participants. Twenty-four students (nine male, mean age 22.41 years, *SD* 1.19) from Hiroshima University participated in the experiment in return for 500 yen (US$4.40). All were native speakers and readers of Japanese and had normal or corrected-to-normal vision.

Design. The experiment was a masked priming lexical-decision task. The targets were preceded by one of four types of primes: (a) katakana identity, (b) hiragana identity, (c) mixed kana identity, and (d) control. The dependent variables were response latency and error rate.

Materials. The critical stimuli were identical to that of Experiment 1, except that the target words were presented in hiragana.

Apparatus and procedure. The apparatus and the general procedure were identical to Experiment 1.

Results and Discussion

The general analysis procedure was identical to Experiment 1. Correct RTs and error rates for the word targets were analyzed using LME modeling with subjects (24) and items (120) as crossed random factors and as the fixed effect factors frequency (log transformed) and Prime type, referenced to the mixed kana condition. In the analysis of RTs, error trials (237 trials, 8.23%) were excluded, and the RTs were log transformed to meet the distributional assumptions of LME. Upon the inspection of Q-Q plot, RTs slower than 1,500 ms were excluded as outliers, resulting in 2,635 data points for analysis.

RT. The mean correct RT and error rates for Experiment 2 are shown in Table 1.

The final statistical model we report here is: logRT ~ logfreq + Primetype + (1 | word) + (1 + primetype | subj), with the Primetype referenced to the mixed kana identity prime condition. The effect of frequency was significant, $t = -4.909$, $B = -.026$, $SE = .005$, $p < .001$. The control condition was significantly slower than the mixed kana identity prime condition, $t = 7.272$, $B = .096$, $SE = .001$, $p < .001$, indicating that the mixed kana prime produced a significant priming effect. As in Experiment 1, the katakana identity prime condition was significantly faster, $t = -5.201$, $B = -.049$, $SE = .009$, $p < .001$, indicating that the priming effect produced by the mixed kana identity prime was smaller than that produced by the katakana identity prime. The Bayes factor for the difference between the katakana identity and mixed kana identity prime was 49,765 against the null model, indicating very strong support for the reduction in the size of priming effect produced by the latter. The hiragana identity prime condition did not differ significantly from the mixed kana identity condition, $t = -0.585$, $B = -.006$, $SE = .011$, $p = .563$. The Bayes factor for this comparison was about 13 in favor of the null model. Referenced to the hiragana identity prime condition, the control condition was significantly slower, $t = 7.222$, $B = .103$, $SE = .014$, $p < .001$, and the katakana identity prime condition was significantly faster, $t = -3.623$, $B = -.043$, $SE = .012$, $p <$

[3] The dash character indicates that the preceding vowel should be lengthened. The point that it is used only in katakana notation was also noted by Perea et al. (2017, footnote 2, p. 1141).

.001. The Bayes factor for the comparison between the katakana and hiragana identity prime condition was 510 against the null model, indicating strong evidence for the reduction in priming in the hiragana identity prime condition. In sum, the pattern of priming was identical to that of Experiment 1, indicating katakana identity < hiragana identity = mixed-kana identity < control.

Error rate. As in Experiment 1, error rates were analyzed with generalized linear mixed-effects model with subjects and stimuli as crossed random factors, using the logit function appropriate for categorical variables. The statistical model we report here is: error ~ logfreq + Primetype + (1 | word) + (1| subj), with the Primetype referenced to the mixed kana identity prime condition. Frequency significantly reduced errors, Z = −3.958, $B = -.34127$, $SE = .8622$, $p < .001$. Relative to the mixed kana identity prime, the control condition was significantly more error prone, Z = 2.549, $B = .50689$, $SE = .19882$, $p < .001$; and the katakana identity prime condition was significantly less error prone, Z = −3.060, $B = -.75734$, $SE = .24747$, $p < .001$. The same model referenced to the hiragana identity prime condition revealed that relative to this condition, the katakana identity prime condition was significantly less error prone, Z = −3.803, B −.92548, $SE = .24334$, $p < .001$.

Experiment 2 presented the loanword targets in hiragana. This meant that the hiragana identity prime now had the same form as the target (e.g., れすとらん - れすとらん), and the katakana identity prime had no form overlap (レ ストラン - れすとらん). Despite this, the pattern of priming remained unchanged from Experiment 1: The katakana prime produced the greatest amount of priming, and the hiragana and mixed-kana primes did not differ from each other, and both produced smaller priming than the katakana prime. The results thus indicate that the size of priming was not determined by the number of characters shared between the prime and target. Instead the results are consistent with the orthographic cue hypothesis, which suggests that for loanwords in Japanese the orthographic format (i.e., it is written in katakana) is represented in the stored orthographic representation, and explains the reduced priming effect with the mixed-kana and hiragana identity prime in terms of a lack of match with this orthographic representation.

General Discussion

The emerging literature on Japanese word recognition indicated a puzzling contradiction. On the one hand, robust cross-kana priming effects have been observed in an allograph match task, suggesting that katakana and hiragana characters map onto a shared abstract character identity (Kinoshita et al., 2019, with native Japanese readers; Schubert et al., 2018 with adult L2 learners of Japanese). On the other hand, in lexical decision with loanwords written in katakana as targets, Perea et al. (2017) found that relative to the katakana identity prime, priming is reduced with the mixed-kana identity prime, a finding interpreted by the authors as indicating that "characters in the two scripts do not share abstract character-level representations" (p. 1145). The aim of the present study was to resolve this puzzle and develop a coherent account of the data. In so doing, we highlight the pitfalls of letter-word equivalence assumption (Bowers et al., 1998), that is, the practice of using the results obtained with word (or pseudoword) stimuli to draw implications for letters (scripts).

Kinoshita et al. (2019) have earlier suggested that the reduced priming observed by Perea et al. (2017) with their mixed-kana prime may have been due to the unfamiliar orthographic format: Specifically, borrowing from the orthographic cue hypothesis (Peressotti et al., 2003) originally suggested for proper names written in the Roman alphabet (e.g., Rome, Elizabeth), it was suggested that for loanwords in Japanese the allographic format (katakana) may be specified in the orthographic representation stored in their long-term knowledge. Thus, just as the recognition of proper names is hampered when presented in an unfamiliar orthographic format (e.g., rome, elizabeth), lexical access of Japanese loanword is hampered when presented in the mixed-kana format. This interpretation found support in Experiment 1, which showed that both the mixed-kana prime and the hiragana prime—which is also orthographically unfamiliar at the word level—produced reduced priming relative to the katakana prime. Experiment 2 replicated this pattern with targets presented in the hiragana format, thus ruling out the possibility that the greater priming observed with the katakana prime in Experiment 1 was due to a greater number of characters shared between the prime and target. In sum, the pattern of priming effects produced by the identity primes in various kana formats— katakana, hiragana, and mixed-kana—is entirely consistent with the account that assumes that for foreign loanwords the katakana format is specified in the orthographic representation stored in the Japanese reader's mental lexicon.

Difference Between Writing Systems or Types of Words?

As noted by Perea et al. (2017), the pattern of priming effects observed with the Japanese katakana words is quite different from those that has been observed with languages written in the Roman alphabet. Here for target words written in uppercase, identity-priming effects of the same size are observed for the identity primes written in uppercase (e.g., BEARD-BEARD), lowercase (beard-BEARD), or mixed case (BeArD-BEARD; e.g., Forster, 1998; Jacobs et al., 1995; Perea et al., 2015). Although Perea et al. (2017) attributed these differences to "the idiosyncrasies of each *writing system* (emphasis ours)" (p.1145), a more parsimonious explanation might be that the differences reflect the different type of words, more specifically, the orthographic format convention associated with different types of words. Unlike the Japanese loanwords, common words in English and other languages written in the Roman alphabet are not tied to a specific allographic format— common words are written and encountered in uppercase letters (e.g., BEARD) as well as lowercase (beard). To put it another way, for these words, the orthographic representation stored in the mental lexicon of the reader does not specify the allographic variant used to write these words but is represented in terms of abstract letter identities.

In contrast to the common words, even within English and other languages written in the Roman alphabet, there are words that are associated with a specific orthographic format. Proper names like personal names (e.g., Elizabeth, Maria) and geographical place names (e.g., Rome, America) are one such example, and Peressotti et al. (2003) have shown that lexical decision to these words are slowed when presented in an unusual orthographic format (e.g., rome, elizabeth). Brand names like IKEA and adidas are another example. It is of interest to note that Perea, Jiménez, Talero, and

Lo′pez-Cañada (2015) appealed to the orthographic cue hypothesis to explain the result they observed with brand names. In a brand name decision task, for brand names tthat are usually written in the lowercase format (e.g., adidas), they reported that priming was reduced when the masked identity prime was in the unfamiliar format (ADIDAS-adidas) than in a familiar format (e.g., adidas-adidas).[4] A more recent study by Martin and Davis (2019) did not replicate this specific result but found a main effect of prime case with participants responding fastest to all brands when preceded by identity primes presented in titlecase (e.g., Adidas-adidas, Audi-Audi, Honda-HONDA). The authors noted that their interpretation that participants exploit title case format to make brand name decisions is similar in many respects to the orthographic cue hypothesis proposed by Peressotti et al. (2003). These studies indicate that the effects of orthographic format of masked primes are not limited to Japanese words written in katakana but are found also with certain types of words written in the Roman alphabet, indicating that the difference in the pattern of priming observed with common words in English (e.g., beard/BEARD) and loan-words in Japanese may be better explained in terms of the difference in the types of words, rather than difference in languages or writing systems.

Another type of words written in a specific allographic format in the Roman alphabet is abbreviations (e.g., CIA, IBM). It should be noted in this context that Brysbaert, Speybroeck, and Vanderelst (2009) used abbreviations as masked primes in lexical decision and did not find effects of case format: Uppercase, lowercase, or mixed case primes (e.g., BMX, bmx, bMx) produced the same-size effect, in apparent contradiction to what is expected from our orthographic cue account. What should be noted, however, is that the type of priming investigated here was not identity priming but associative priming (e.g., BMX-BIKE vs. CIA-BIKE). Associative priming effect is not an orthographic effect—BMX and BIKE are related semantically, not orthographically. This contrasts with the identity priming effect (e.g., beard-BEARD vs. child-BEARD) which is primarily orthographic (see Kinoshita, Gayed, & Norris, 2018). Consistent with this analysis, an experiment recently completed in our laboratory (Kinoshita & Whiting, 2019) showed that the case format of the abbreviation prime did modulate the identity priming effect (e.g., BMX-BMX < bmx-BMX), but it had no effect in an associative priming experiment (BMX-BIKE = bmx-BIKE).

Conclusion

The kana format of masked identity prime has yielded mixed findings that have led to contradictory claims regarding whether hiragana and katakana share abstract character representations. Here we have pointed out that it is not prudent to interpret results obtained with word stimuli as reflecting the nature of letter/character representations (the letter-word equivalence assumption, Bowers et al., 1998), and presented a unified account of the seemingly contradictory findings based on the assumption that the katakana format is specified in the orthographic description of Japanese loanwords in the mental lexicon. Our view is that both the notion of ALIs shared by the allographic variants and the notion that allographic format serves as a cue to the linguistic status of certain classes of words (proper names in Italian and English, nouns in German) as originally suggested by the orthographic cue hypothesis (Peressotti et al., 2003) are applicable across languages and writing systems, and we hope this view will guide future work in visual word/letter recognition.

[4] Martin et al. (2019) noted that what Perea et al. (2015) described as "brand names which are usually written in the lowercase format" in fact contained both brand names like adidas and brand names usually written in title case such as Ford, Audi, and Lipton.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511801686

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2018). Lme4: Linear mixed-effects models using 'Eigen' and S4 (Version 1.1–17) [Computer software]. Retrieved from https://cran.r-project.org/package=lme4

Bigsby, P. (1988). The visual processor module and normal adult readers. *British Journal of Psychology, 79,* 455–469. http://dx.doi.org/10.1111/j.2044-8295.1988.tb02746.x

Bowers, J. S., Vigliocco, G., & Haan, R. (1998). Orthographic, phonological, and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1705–1719. http://dx.doi.org/10.1037/0096-1523.24.6.1705

Brysbaert, M., Speybroeck, S., & Vanderelst, D. (2009). Is there room for the BBC in the mental lexicon? On the recognition of acronyms. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 62,* 1832–1842. http://dx.doi.org/10.1080/17470210802585471

Carreiras, M., Perea, M., & Abu Mallouh, R. (2012). Priming of abstract letter representations may be universal: The case of Arabic. *Psychonomic Bulletin & Review, 19,* 685–690. http://dx.doi.org/10.3758/s13423-012-0260-8

Chikamatsu, N., Yokoyama, S., Nozaki, H., Long, E., & Fukuda, S. (2000). A Japanese logographic character frequency list for cognitive science research. *Behavior Research Methods, Instruments, & Computers, 32,* 482–500. http://dx.doi.org/10.3758/BF03200819

Coltheart, M. (1981). Disorders of reading and their implications for models of normal reading. *Visible Language*. Retrieved from http://psycnet.apa.org/PsycINFO/1982-29559-001

Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J.-B., Le Bihan, D., & Cohen, L. (2004). Letter binding and invariant recognition of masked words: Behavioral and neuroimaging evidence. *Psychological Science, 15,* 307–313. http://dx.doi.org/10.1111/j.0956-7976.2004.00674.x

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5,* 781. http://dx.doi.org/10.3389/fpsyg.2014.00781

Forster, K. I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research, 27,* 203–233. http://dx.doi.org/10.1023/A:1023202116609

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116–124. http://dx.doi.org/10.3758/BF03195503

Jacobs, A. M., Grainger, J., & Ferrand, L. (1995). The incremental priming technique: A method for determining within-condition priming effects. *Perception & Psychophysics, 57,* 1101–1110. http://dx.doi.org/10.3758/BF03208367

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, New York: Oxford University Press, Clarendon Press.

Kinoshita, S., Gayed, M., & Norris, D. (2018). Orthographic and phonological priming effects in the same-different task. *Journal of Experimental Psychology: Human Perception and Performance, 44,* 1661–1671. http://dx.doi.org/10.1037/xhp0000548

Kinoshita, S., & Kaplan, L. (2008). Priming of abstract letter identities in the letter match task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 61,* 1873–1885. http://dx.doi.org/10.1080/17470210701781114

Kinoshita, S., Schubert, T., & Verdonschot, R. G. (2019). Allograph priming is based on abstract letter identities: Evidence from Japanese kana. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 183–190. http://dx.doi.org/10.1037/xlm0000563

Kinoshita, S., & Whiting, D. (2019, November). WTF? Prime case effect for abbreviations. Paper presented at the 60th annual meeting of the Psychonomic Society, Montreal, Canada.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). lmerTest: Tests in linear mixed effects models (Version 3.0-1) [Computer software]. Retrieved from http://CRAN.R-project.org/package=lmerTest

Martin, N., & Davis, C. (2019). Evidence from masked-priming that initial identification of brand names is via abstract letter identities. *British Journal of Psychology*. Advance online publication. http://dx.doi.org/10.1111/bjop.12362

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs (R package, version 0.9.12-4.1) [Computer software]. Retrieved from http://CRAN.R-project.org/package=BayesFactor

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General, 137,* 434–455. http://dx.doi.org/10.1037/a0012799

Perea, M., Jiménez, M., Talero, F., & Lo′pez-Cañada, S. (2015). Letter-case information and the identification of brand names. *British Journal of Psychology, 106,* 162–173. http://dx.doi.org/10.1111/bjop.12071

Perea, M., Nakayama, M., & Lupker, S. J. (2017). Alternating-script priming in Japanese: Are Katakana and Hiragana characters interchangeable? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1140–1146. http://dx.doi.org/10.1037/xlm0000365

Perea, M., Vergara-Martínez, M., & Gomez, P. (2015). Resolving the locus of cAsE aLtErNaTiOn effects in visual word recognition: Evidence from masked priming. *Cognition, 142,* 39–43. http://dx.doi.org/10.1016/j.cognition.2015.05.007

Peressotti, F., Cubelli, R., & Job, R. (2003). On recognizing proper names: The orthographic cue hypothesis. *Cognitive Psychology, 47,* 87–116. http://dx.doi.org/10.1016/S0010-0285(03)00004-5

Polk, T. A., Lacey, H. P., Nelson, J. K., Demiralp, E., Newman, L. I., Krauss, D. A., . . . Farah, M. J. (2009). The development of abstract letter representations for reading: Evidence for the role of context. *Cognitive Neuropsychology, 26,* 70–90. http://dx.doi.org/10.1080/02643290802618757

Pylkkänen, L., & Okano, K. (2010). The nature of abstract orthographic codes: Evidence from masked priming and magnetoencephalography. *PLoS ONE, 5,* e10793. http://dx.doi.org/10.1371/journal.pone.0010793

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Schubert, T., Gawthrop, R., & Kinoshita, S. (2018). Evidence for cross-script abstract identities in learners of Japanese kana. *Memory & Cognition, 46,* 1010–1021. http://dx.doi.org/10.3758/s13421-018-0818-4

Schubert, T., & McCloskey, M. (2013). Prelexical representations and processes in reading: Evidence from acquired dyslexia. *Cognitive Neuropsychology, 30,* 360–395. http://dx.doi.org/10.1080/02643294.2014.880677

Tamaoka, K., Makioka, S., Sanders, S., & Verdonschot, R. G. (2017). A new interactive online database for psychological and linguistic research on Japanese kanji and their compound words. *Psychological Research, 81,* 696–708. http://dx.doi.org/10.1007/s00426-016-0764-3

Received April 29, 2019
Revision received July 8, 2019
Accepted July 9, 2019 ■