Article:

Burt, J.S., McFarlane, K.A., Kelly, S.J. et al. (3 more authors) (2018) Brand name confusion: subjective and objective measures of orthographic similarity. Journal of Experimental Psychology: Applied, 23 (3). pp. 320-335. ISSN 1076-898X

https://doi.org/10.1037/xap0000127

Running Head: Orthographic similarity ratings

Brandname confusion: Subjective and objective measures of orthographic similarity.

Jennifer S. Burt

Kimberley A. McFarlane

Sarah J. Kelly

Michael S. Humphreys

University of Queensland,

Brisbane, Australia

Kimberlee Weatherall

University of Sydney

Sydney, Australia

Robert G. Burrell

University of Sheffield

Sheffield, United Kingdom

Corresponding author:

Jennifer S. Burt

School of Psychology

University of Queensland

QLD 4072

Australia

email: j.burt@psy.uq.edu.au

Phone: +61 7 3365-6338

Facsimile: +61 7 3365-4466

Abstract

Determining brandname similarity is vital in areas of trademark registration and brand confusion. Students rated the orthographic (spelling) similarity of word pairs (Experiments 1, 2, and 4) and brandname pairs (Experiment 5). Similarity ratings were consistently higher when words shared beginnings rather than endings, whereas shared pronunciation of the stressed vowel had small and less consistent effects on ratings. In Experiment 3 a behavioral task confirmed the similarity of shared beginnings in lexical processing. Specifically, in a task requiring participants to decide whether two words presented in the clear (a probe and a later target) were the same or different, a masked prime word preceding the target shortened response latencies if it shared its initial three letters with the target. The ratings of students for word and brandname pairs were strongly predicted by metrics of orthographic similarity from the visual word identification literature based on the number of shared letters and their relative positions. The results indicate a potential use for orthographic metrics in brandname registration and trademark law.

The present studies were motivated by issues in brandname similarity, and in particular by our belief that the insights and tools of researchers in visual word identification have under-realised potential to contribute to decision-making in both law (especially trademark law) and marketing.

In trademark law, it is often necessary to judge the degree of similarity between two words. Examiners working within government intellectual property offices considering whether to accept a trademark for registration may need to decide whether it is too similar to an existing registered mark in the same or a similar product category. In legal disputes under trademark or consumer protection law, a court may need to decide whether an alleged infringer's name is too similar to a name used or registered by another. In each of these legal scenarios, the legal question in most countries turns on whether two brandnames are 'confusingly' or 'deceptively' similar. The answers to these similarity questions depend in turn on an assessment of how consumers of the products will perceive the two brandnames - the assumption being that perceived similarity will cause consumers to be confused about product origin. As a result, consumers may, for example, purchase the allegedly infringing goods in the mistaken belief that they are the goods of the mark owner. Alternatively, they may draw an association between them; in particular, that the goods have been produced under licence from the trademark owner (such that consumers will be more willing to try the allegedly infringing goods and/or blame the trademark owner if the goods prove unsatisfactory).

Understanding the degree to which brandnames are similar is also important for the creation, protection and growth of brands. In a shopping context, brandnames will be searched or browsed in the on-line environment or in a store. In many cases shoppers may search for a specific familiar brand. Nevertheless, because of time pressure, they may devote only cursory processing to brandnames (Chandon, Hutchinson, Bradlow, & Young, 2009; Van der Lans, Pieters, & Wedel, 2008). Given this situation, copycat strategies – whereby newer entrant brands imitate packaging or other features of existing original or leading brands – are common (e.g., Van Horen & Pieters, 2012; Walsh, Mitchell, Kilian, & Miller, 2010). An example involving brandnames (Van Horen & Pieters, 2012) is the name *Ozemite,* which may be perceived as similar to the established Australian brandname *Vegemit*e. Copycat strategies are a growing commercial strategy with potentially adverse effects on established brands. For both defenders and critics of such strategies, similarity matters. Consumers will only understand a product to be a substitute if consumers associate the copycat product with the original; producers would not need to worry about confusion if consumers would not respond to the products as being at least 'similar'.

Brand owners have a right to prevent the use of similar trade indicia, including brandnames. A vital question is how trademark examiners and judges in legal disputes decide when two names are similar enough to risk confusion. Surprisingly, a decision is often made by a single observer about the perceived visual similarity. For example, in legal disputes, a judge may decide on name similarity without reference to any objective measure, and without reference to any subjective measure whose reliability and validity can be defended scientifically. Empirical evidence adduced through surveys may be rejected in legal cases on the basis of being non-representative, leading, not ecologically valid or having samples of insufficient size (Dinwoodie & Gangjee, 2015; Huang, Weatherall, & Webster, 2012). This reluctance is, in some cases, unfounded. Psychological research shows that human judgments are subject to unreliability when only a single observer is used, but large increases in reliability can be achieved by averaging over the judgments of a relatively small group of observers. Contrary to lawyers' assumptions, when judgments about name similarity are made, there is no basis for claiming that the vast majority of English speakers will differ in their judgments as a function of demographic variables, or that very large samples of observers are required.

In the present studies, groups of approximately twenty university students made ratings about the similarity (likely confusion in reading) between two words or two brandnames. Participants' attention was drawn to word spelling; no mention was made of word meaning. There were two broad aims.

The first was to make a targeted assessment of factors that might affect similarity. Of particular interest was the assumption commonly made in Australian and UK law that word beginnings have a substantial impact on similarity, with words sharing their beginnings being perceived as more similar than words overlapping in non-initial components: *London Lubricants* (1925) 42 RPC 264. An additional question was whether similarity of pronunciation would impact similarity judgments about visually presented words,[1] and the impact of word length. Courts regularly consider all these aspects of similarity, but with little by way of empirical support for judicial assumptions about what factors are most important in judgments of similarity, or how these factors interact (Burrell & Handler, 2016, pp. 206-208).

The second aim was to assess predictors of students' mean similarity ratings, with a view to finding the best metric for making reliable estimates of the similarity judgments of groups of people. If such a metric can be found, the efficiency of similarity checking could be improved in trademark registration. When registration of a new trademark is sought, in many (but not all) countries examiners search the trademark register – a very large database of existing and past trademark registrations – for marks that might be too similar. This generates long lists of possibly similar marks that examiners must then narrow down to those most likely to be confusingly similar. A metric could facilitate this search by producing brandnames to be considered in detailed similarity assessments or assisting in the ranking of initial search results. We recognise that in making a final decision on registrability, a trademark examiner must take into account a number of factors beyond physical name similarity, including the nature of the product categories, semantic connotations, and the implications for everyday language use.

To accomplish these aims we used a metric for stimulus selection and controlling overall orthographic similarity in item subsets. In the final two studies we added two theory-based metrics from the word reading literature to ascertain the best predictor for legal application. We

[1] In legal decisions, judges pay attention to similarity in pronunciation of words: *Wingate Marketing Pty Ltd v Levi Strauss & Co* (1994) 49 FCR 89. However visual similarity is more important for goods that will be selected from a shelf or otherwise visually presented: *Taiwan Yamani Inc v Giorgio Armani SpA* (1989) 17 IPR 92.

also evaluated a phonological measure as a predictor in case similarity effects were driven by the pronunciation similarity of orthographically similar words.

The choice of a metric for word pair selection was governed by our requirement for a validated measure that would yield a range of similarities. Recent work in visual word reading with words covering a large length range (Yarkoni, Balota, & Yap, 2008) has empirically validated a metric from computer science as a predictor of word reading efficiency. This metric, the Orthographic Levenshtein distance (OLD) was used for item selection in all experiments, and as a predictor in our final two studies. It is based on the number of operations (insertions, deletions, substitutions) required to turn one letter string into another, with OLD increasing as orthographic similarity decreases. We used the Damerau variation, which unlike the traditional metric, counts the swapping of two adjacent letters as one operation (Keller, 2014).

For the comparison of predictors we added two orthographic metrics from the visual word identification literature, namely the unweighted and end-weighted orthographic match values (Davis, 2007) from the Spatial Coding Model of visual word identification (Davis, 2010). These metrics were derived within a model of visual word identification whose central focus is the encoding of letter order by the lexical processing system, and the effects of orthographic similarity on word identification. The model codes the spatial position of letters in a letter string, going from one end of the string to the other, and then assigns each letter a position with some uncertainty, represented by a distribution of activation that falls as the distance from the actual position increases. A match value between two letter strings is calculated when an input string is matched with its internal memory representation to achieve word identification. Like OLD, the match value is sensitive to the number of shared letters and their positions in the two letter strings. Although there are other theories of letter position coding (Grainger & Van Heuven, 2004; Whitney, 2001), the Davis model was chosen because it is a well-developed model of word identification, has empirical support (Burt & Duncum, 2017), and has an accessible stand-alone calculator for computing a similarity between 0 and 1 for a pair of letter strings. Our primary goal was to find robust, practically useful, subjective and objective measures of orthographic similarity that apply to words in general, rather than the short (mainly one-syllable) words typically used by reading researchers.

The development of measures of orthographic similarity connects with contemporary issues in research in visual word identification. The coding of letter position in to-be-read words is currently a focus of attention in theories of word reading (Davis, 2010; Grainger & Van Heuven, 2004; Norris & Kinoshita, 2012). Traditional models of word reading have letters coded in position-specific slots, so that words like *caterpillar* and *capillary* or *cart* and *arts* do not activate each other's memory representations because their shared letters occupy different slots (Morton, 1969). More recently it has become clear that strict position-specific letter coding does not capture the behaviour of readers. For example, it is evident that reading can be successful (albeit slower, Rayner, White, Johnson, & Liversedge, 2006) when the internal letters of a printed word are re-arranged. This fact was demonstrated in the so-called "Cambridge email", according to which "it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae". Together with laboratory results (Lupker, Perea, & Davis, 2008; Perea & Lupker, 2004), this result indicates that a printed word with its internal letter-order perturbed can be successfully matched with the word's orthographic representation stored in a reader's memory. We expect subjective ratings to provide evidence that converges with the findings from behavioral tasks, on the grounds that participants' judgments of word similarity reflect the structure of their language, their experience with the written language, and how these factors have shaped participants' lexical processing systems.

Although the primary interest was in brandnames, Experiments 1-4 involved English words because they allowed better control over item selection. Brandnames in Australia generally accord with the structure of English words, and we expected word results to generalise to brandnames. Experiments 1 and 2 were designed to provide a controlled assessment of shared orthographic or phonological features in word sets that were matched on overall similarity (as indexed by OLD) and on other relevant variables. For ease of exposition we report only analyses by participants, but we note relevant information from the analyses with items as the random effect. The results were assessed in a new behavioral task in Experiment 3. In Experiments 4 and 5, a large range of item lengths and OLD values were sampled in order to make a generalizable assessment of the predictors of similarity ratings. For predictive analyses we divided OLD by the length of the longer pair member to produce a value between 0 and 1, termed here OLDscaled. Otherwise the maximum OLD would depend upon the length of the longer pair member and the role of OLD could not be disentangled from the effects of item length.

Experiment 1

The first experiment required participants to make judgments on a scale of 1 to 4 about the orthographic similarity of two words presented side by side in the center of a computer display. Each pair consisted of a target word of 6 letters in length, and a comparison word of 5, 6, or 7 letters. Each target was seen twice by each participant, once with a similar word and once with an unrelated (dissimilar) word. The similar targets were matched at a moderately high similarity (OLD = 2). In order to make the rating task meaningful, fillers with OLD distances of 1 and 3 were added. The fillers and similar critical items were compared in order to assess whether participants' ratings did vary with OLD.

The primary aim was to assess the effect of phonological similarity. Vowels are particularly important in phonological effects in lexical tasks (such as priming effects by phonologically similar words that precede a target word), perhaps because there is more ambiguity about the pronunciation of vowels than consonants (Berent & Perfetti, 1995; Treiman, Kessler, Zevin, Bick, & Davis, 2006). The strongest priming effects have generally been found for words sharing the rime (vowel plus final consonants) of a one-syllable word, as in the pair *mate-rate* (Taraban & McClelland, 1987). In the present studies the aim was to make assessments that applied to words in general, rather than to the subset of one-syllable words. Consequently, we focused on whether word pairs that shared or did not share the stressed vowel. It is important to note that the term vowel refers to the pronunciation; the orthography is not necessarily the same for identical vowels (cf. *oe* and *ow* in *hoe* and *low*). We matched the words in each pair on orthographic similarity to address the confound of orthographic and phonological similarity.

Filler pairs were added to provide a range of similarity as indexed by OLD.

Method

Participants. Twenty-four introductory Psychology students participated for course credit.

Materials and Design.

The critical items were 104 target words of length 6 letters and mean frequency 31 per million (range 17 - 50; Kilgarriff, 1995). Each was paired with a similar word of 5, 6, or 7 letters that had an Orthographic Levenshtein Distance (OLD) of 2; that is, 2 letter changes were required to change the target into the comparison word. The OLDscaled measure ranged from .29 to .33 for the similar pairs in the critical sets in this experiment and in Experiment 2. There were 40 targets with 5-letter comparison words and 40 with 7-letter comparison words. The remaining 24 had 6-letter comparison words. For each similar comparison word there was a length-matched unrelated word sharing no more than 2 letters in position with the target. The mean OLDscaled value for unrelated pairs was .91. The similar pairs are shown in Appendix A.

In addition there were 2 sets of 20 filler pairs, one with an OLD of 3 (e.g., *whisky-thinks*) and another with an OLD of 1 (e.g., *lively-lovely*). All filler words were 6 letters in length and had a mean frequency of 31 per million (range in the British National Corpus, Kilgarriff, 1995). Thus in the total item set, 44% of the trials had equal-length pair members.

For the similar comparison words at each length and their critical targets, and also for the filler pairs, half of the pairs shared the stressed vowel in their pronunciation: for example, *relate-replace*. The remaining pairs did not share the stressed vowel: for example, *wished-sighed*.

Two counterbalanced lists were constructed that were identical except for the left-right position of the target in the word pairs. Each list contained all 40 filler pairs and the 104 critical targets with both of their comparison words. Thus, each target was seen twice by each participant, once with its similar comparison word and once with its unrelated control. The position of the target (left vs. right) was different for the two target presentations, and the position in similar and unrelated pairs was reversed from list 1 to list 2. The trial sequence was randomised and seven practice trials covering a range of similarities were added to the beginning of each list.

Procedure. An E-prime program (Schneider, Eschman, & Zuccolutto, 2002) presented words and collected responses. The text was displayed in 18-point courier font against a dark blue background. On each trial a ready signal appeared in white (+++) for 250 ms, and then the two words were presented in white side by side in the centre of the screen at a separation of about 4 cm. Underneath the word pair was a reminder of the scale in lime green font. The digits 1, 2, 3 and 4 were arrayed from left to right across the screen, with the label *Not Similar* beneath the 1, and *Very Similar* beneath the 4. Participants typed in a number from 1 to 4, and the word pair and scale was cleared from the screen and a 2 sec inter-trial interval began.

The instructions asked participants to make a judgment (based on the letters and letter order) about the words' similarity, in the sense of their being easily confused in reading. They were given the examples *salt-slat*, *silk-slat* and *book-slat* as ranging from high to low similarity. They were asked to make an intuitive judgment and guess if not sure.

The 248 trials plus 7 practice trials were present in 4 blocks of 51 trials, with a self-paced rest between blocks.

Results and Discussion

Mean ratings were submitted to two ANOVAs to assess OLD variation and the effect of vowel match. Effects that were significant by participants were also significant by items, unless indicated otherwise. The first ANOVA assessed the effect of OLD in the OLD-1 and OLD-3 fillers plus the similar pairs (OLD-2) from the critical set. A one-way ANOVA for

OLD (1 vs. 2 vs. 3), collapsing over vowel match, revealed a robust effect, $F(2, 46) = 345.58$, $\eta_P^2 = .94$, with mean ratings of 3.30, 2.43 and 1.92 for OLD 1, 2 and 3 respectively.

The second ANOVA assessed the effects of Vowel match (match vs. mismatch) x Similarity (similar vs. unrelated control) x Comparison length (5 vs. 6 vs. 7 letters) on the ratings for the critical targets in their similar and unrelated pairs. The mean ratings are shown in Figure 1 as a function of target length. As can be seen in the figure, the controls received ratings close to the minimum score of 1. There was a robust effect of similarity, confirming that OLD-2 pairs were rated as substantially more similar than the unrelated controls, $F(1, 23) = 494.36$, $\eta_P^2 = .96$. With respect to length, Figure 1 shows that 5-letter comparisons tended to produce lower similarity ratings than the other pairs, with length effects somewhat different over similar and control items. The Comparison length x Similarity interaction was only marginally significant in the items analysis and will not be discussed further.

The contribution of vowel match is best captured in the three way interaction of Vowel x Similarity x Comparison length, $F(1, 23) = 23.93$, $\eta_P^2 = .51$ ($p = .1$ by items). As is evident from the figure, vowel match only had an effect when both the target and its comparison word were 6-letters long. There was a higher similarity rating for the vowel-match similar pairs than the vowel-mismatch similar pairs when the length of both words was 6 letters, $F(1, 23) = 27.92$. This simple effect was also significant in the items analysis ($p = .01$).

Insert Figure 1 about here

In summary, the results of Experiment 1 confirmed that considerable variance in participants' ratings was captured by the OLD metric. This finding is in line with recent findings that a word's OLD distances from other words explains variance in response latencies in the lexical decision task (LDT, Yarkoni et al., 2008). Most importantly, the result confirms that objective metrics have the potential to capture consumers' perceptions of the relative similarity of brandname pairs. A match in the stressed vowel for similar pairs increased similarity ratings for critical pairs somewhat, but only when the pair members were matched in length. Consequently, phonological similarity as indexed by a stressed-vowel match had modest and constrained effects on similarity ratings.

Experiment 2

Experiment 2 was similar in design to Experiment 1, but the focus of interest was on beginning vs. end overlap for similar pairs. As noted previously, one aim of the present studies was to evaluate the validity of assumptions about similarity made in the law. A legal assumption is that perceived similarity of brandnames is enhanced by a beginning overlap in the names (Burrell & Handler, 2016). Although this assumption is shared in the lexical processing literature, there is to our knowledge no evidence on the effects of beginning overlap on subjective ratings of similarity.

New similar pairs (OLD-2) shared the first 3 letters or the last 3 letters. The match on OLD ensures that it is the beginning overlap rather than overall similarity that is important. In addition, the fillers and a small subset of the vowel-match and control items were taken from Experiment 1. The effects of a vowel match were not significant in this smaller item set, so the results for these items are not reported.

Method

Participants. A new sample of twenty-four introductory Psychology students participated for course credit.

Materials and Design. Ten targets and comparison words were taken from the each of the vowel match and mismatch sets of Experiment 1. Each set of 10 had two targets paired with equal length comparison words and 4 targets paired with 5 and 7-letter comparison words. The Experiment 1 filler pairs were also included. They produced similar results to Experiment 1 and the analysis over OLD values is not reported.

The new items were 84 6-letter targets that had similar (OLD-2) vs. unrelated comparison words. The similar words differed in beginning vs. end overlap with the target. That is, the target and comparison shared the initial three letters, for example, *waited-waist*, (N= 42); or the three final letters, for example, *remove-prove* (N = 42, see Appendix B). Within each set of 42 a third of the targets were allocated to each of the three comparison-word lengths (5, 6 and 7 letters). Each target had a similar and unrelated comparison word, and participants saw each target once in each pairing (as in Experiment 1). The OLDscaled mean for unrelated words was .90. The trial sequence was randomised and 7 practice trials were given, making a total of 255 trials.

Procedure. The procedure was as in Experiment 1.

Results and Discussion

The analysis examined the effect of beginning vs. end overlap in the new OLD-2 pairs and their unrelated controls. An Overlap (beginning vs. end) x Comparison length (5 vs. 6. vs. 7 letters) x Similarity (similar vs. unrelated control) ANOVA was conducted on ratings. All main effects and interactions were significant by participants; only key effects will be reported (see Figure 2). As before, there was a large difference between similar pairs and their controls, $p < .001$. There was a main effect of length, with higher similarity ratings for equal-length pairs, $F(1, 23) = 22.43$, $\eta_P^2 = .49$. This was especially so for similar pairs, although the Length x Similarity interaction fell short of significance in the items analysis. Of primary interest, the Overlap x Similarity interaction was significant, confirming a larger advantage for similar pairs over controls for beginning overlap than for end overlap, $F(1, 23) = 25.53$, $\eta_P^2 = .52$. The three-way interaction of Overlap x Comparison length x Similarity was not significant by items ($p = .34$).

Insert Figure 2 about here

In a smaller item set we failed to see an effect of a vowel match (analyses not reported). By contrast, in the OLD-1 and OLD-3 fillers, a vowel match produced a small increase in judged similarity (means of 2.45 vs. 2.65). This effect was significant by participants in a post hoc test, $F(1, 23) = 8.64$, $\eta_P^2 = .27$ ($p = .08$ by items). Overall the effect of a vowel match was small and variable, being evident only when pair members were equal length in Experiment 1. With respect to length effects in the two experiments, there was a small decrement in ratings for similar pairs with a 5-letter comparison word.

By contrast with vowel and length effects, the impact of a beginning overlap on similarity ratings was substantial. For OLD-2 word pairs, a beginning overlap produced mean ratings 1.1 points higher than the dissimilar controls, whereas an end overlap produced mean ratings only 0.7 points higher than controls. This result provides support for the legal assumption that beginnings of names are more important than ends in perceptions of similarity. An important qualification is that we can only make this claim about visual presentation, as we have not yet assessed spoken words. Experiment 3 addressed the generality of the beginning effect in a behavioural task.

Experiment 3

If words are judged to be easily confused in reading, then this confusion may be evident in word reading performance. The aim of Experiment 3 was to provide a behavioral validation of subjective judgments. This demonstration would confirm the utility of subjective ratings as an indicator of factors that may affect the behavior of consumers. More generally, a preconscious effect of orthographic similarity would confirm that subjective ratings do reflect perceptions of similarity rather than participants' efforts to respond to the demand characteristics of the rating task. It would also indicate that language users' subjective reports are responsive to language variables that drive pre-conscious lexical processes.

Masked priming paradigms are a useful vehicle for our purposes because the prime word is briefly displayed and sandwiched between forward and backward pattern masks, and thus is not usually available for report by the participant (Forster & Davis, 1984). We used similar and dissimilar control words as masked primes for some of the targets used in Experiment 2 with a view to determining whether a pre-consciously processed similar word would affect response latencies to the target. As outlined below, we used a same-different task rather than the traditional lexical decision task (LDT). In keeping with almost all prior research in masked priming, the prime and target were presented in different letter cases (Forster & Davis, 1984). The purpose is to place the focus on shared orthography (spelling) rather than perceptual similarity effects at the letter level.

Masked priming by orthographically similar word and nonword primes has been extensively investigated (Andrews, 1996; Forster, 1987; Grainger & Ferrand, 1996; Perea & Lupker, 2004), for the most part in the LDT. When primes are words, the results are complex because depending on the frequency of occurrence in text of primes and targets, and the orthographic characteristics of the words and nonwords in the experiment, a word prime sometimes competes with a similar target and delays its recognition (Davis & Lupker, 2006; Nakayama, Sears, & Lupker, 2010). These effects of lexical competition are important for understanding the processes of word identification in reading, but they are not essential for the present goal, which is delineating what words are taken to be orthographically similar by the lexical processing system. A simpler task for examining the latter question is the masked priming version of the same-different task, and this is the task that was used in Experiment 3.

In the same-different task, three letter strings (here, words) are presented successively on each trial, with only two words (the first and last) clearly visible to the participant. The first is displayed for approximately a second and is termed the probe. A brief masked (unseen) prime is then displayed, and finally a target is displayed until a participant responds. The participant's task is to decide whether the probe is the same as the target. If the unseen item, the prime, is orthographically similar to the target, a correct *same* judgment typically is faster than in a control prime condition. Trials in which the probe and target are different are included to make the task work, but no predictions are made about *different* responses. Increases in orthographic similarity of the prime and target tend to increase the priming benefit. As a result, this task has proved useful for asking questions about the effects of orthographic and letter-form similarity in the lexical processing system (Kinoshita & Norris, 2010; Norris &

Kinoshita, 2008). To date there has been no investigation of beginning vs. end overlap or vowel match. In Experiment 3, only equal-length primes and targets were used. Pilot testing indicated some problems with effective prime masking when the primes and targets differed in length.

Method

Participants. A new sample of twenty-four introductory Psychology students participated for course credit.

Materials and Design. The targets were 96 words of mean frequency 31 per million and frequency range 17 to 50 in the British National Corpus (BNC; Kilgarriff, 1995). They included the 32 6-letter-targets used in Experiment 2, and 2 additional sets of 32 words at lengths of 5 and 7 letters. The 32 targets at each length were divided into four sets of 8 for the four item types (beginning vs. end match and vowel match vs. mismatch). Thus, collapsed over the 3 lengths, there were 24 pairs of each type. Item length was not included as a factor in analyses because there were too few observations at each length. Similar and unrelated primes were devised according to item type for each target, with the 6-letter pairs taken from Experiment 2. Primes were matched on length to their targets. Unrelated primes had a mean OLD distance of 5.7 from their targets (maximum distance = 6).

Similar primes in the beginning vs. end match sets shared the 3 initial or final letters of the target, for example, *MODULE-modest* and *RELISH-vanish*, respectively. The beginning match and end match pairs were matched on mean OLD (range of 2 to 4, mean = 2.48), except that by error the mean similarity as indexed by OLD was higher for the beginning-match than end-match pairs for the 7-letter items. Removal of the 7-letter words or the subset of words differing in OLD did not reduce the strength of the effects, so only the analyses for the complete item set are reported. The vowel sets were constructed as in Experiment 2, with the similar prime sharing vs. not sharing the pronunciation of the stressed vowel, without constraint on word beginnings or endings, for example, *PLANE-slate* (same vowel) vs. *ALERT-alarm* (different vowel). All vowel match and mismatch pairs had an OLD of 2. The mean frequency of the similar primes in the BNC (Kilgarriff, 1995) was 25 per million and the primes were approximately matched in frequency over conditions. An additional 96 unrelated words were matched on length and frequency to targets to serve as unrelated probe words on *different* trials. On same trials the probe was the same as the target.

In line with previous studies (Kinoshita & Norris, 2009), each of the 96 targets was seen twice by each participant, once with a probe that matched the target (same trial, *yes* response) and once with a different probe (*different* trial, *no* response). Only the same trials were of direct interest. The two target presentations for each participant had the same prime type (similar vs. unrelated). The targets were cycled through the similar vs. unrelated prime conditions over two counterbalanced lists of 192 trials each. In each list, half of the targets in each length x item-type cell had unrelated primes and the remainder had similar primes. The trial sequence was randomised and eight practice trials were added to the beginning of each list.

Procedure. All items were presented in black 20 point courier font in the center of a white screen. On each trial a ready signal (+++) was displayed for 350 ms, and then the probe word was displayed for 1000 ms in upper case letters. One, two or three 3 hash marks (#) were added to the end of each word to make it 8 characters long. The prime was then displayed in lower case letters for 48 ms, followed by the target in upper case letters, again with hash marks added to give a length of 8 characters. Participants rested their right and left index fingers on the corresponding buttons of a response box, and pressed the right button if the target was the same as the probe word, and the left button if it was different. The response cleared the screen and initiated a 2 sec interval before the next trial. Participants were instructed to respond as quickly and accurately as possible. Response latencies (in ms) were recorded from the onset of the target. The trials were presented in 4 blocks separated by rest breaks.

Results and Discussion

The same vs. different vowel sets and the beginning vs. end overlap sets were analysed separately in line with Experiment 2 and with the separate selection and matching of items for these sets. Recall that as in the ratings experiments, each target was paired with a similar and an unrelated comparison word, with these words appearing as masked primes in the current experiment. Accuracy and latency on *different* trials were examined to check that they did not qualify the interpretation of the same-trial data. Means were similar over conditions for vowel and beginning-end item sets, and there were no effects that were significant by participants and items in the accuracy or latency data for *no* responses.

For the *yes* responses on same trials for the vowel-match set, a Vowel match x Similarity analysis showed no significant effects in error rates, although there was a trend for fewer errors to occur on similar prime trials than on unrelated prime trials (see Table 1). In the latency data, there was only a significant effect of prime similarity, with faster latencies in the similar condition, $F(1, 23) = 11.92$, $\eta_P^2 = .34$. The latency data for all pair types are shown in Figure 3. The benefit of a similar prime was numerically larger for the vowel-match conditions but not statistically so; there was no main or interactive effect of the vowel condition.

Insert Table 1 about here

For the beginning-end sets on same trials (*yes* responses), there were no significant effects in the error data (see Table 1). The Overlap (beginning vs. end) x Similarity (similar vs. unrelated prime) ANOVA on mean latencies showed no main effect of overlap ($F < 1$), a trend (that was significant in the items analysis) for faster latencies on similar- than unrelated-prime trials, $F(1, 23) = 3.46$, $p = .076$, and a significant Overlap x Similarity interaction, $F(1, 23) = 5.32$, $\eta_P^2 = .19$, as shown in Figure 3. The interaction was marginally reliable by items ($p = .06$). Simple effects of prime similarity at each Overlap condition showed a significant priming effect for beginning-overlap pairs, $F(1, 23) = 5.78$, ($p < .001$ in the items analysis), but not for end-overlap pairs ($F < 1$).

Insert Figure 3 about here

The results of Experiment 3 can be summarised simply. As expected, there were effects of prime type on the target latencies for *same* trials but not for *different* trials. Orthographically similar primes decreased *yes* response latencies to targets overall, and the priming benefit was significant within both the beginning-end and vowel sets. In line with the small and variable effects of a vowel match in Experiments 1 and 2, there was no compelling evidence for a larger priming effect when the stressed vowels of the prime and target were the same rather than different. Masked phonological priming effects have been observed in the same-different task (Lupker, Nakayama, & Perea, 2015), so a vowel-match may not produce sufficient phonological similarity. With respect to beginning vs. end overlap, priming benefits were larger for beginning overlap than end overlap pairs. The interactive effect was only marginally significant by items, reflecting high item variability, but the priming effect was robustly significant by items for the beginning-overlap pairs and not significant for the end-overlap pairs. These results are consistent with Experiment 2, which showed similarity ratings to be higher for pairs with beginning than pairs with end matches. The priming effect was not significant for end-overlap pairs taken separately, a result that may reflect the relatively small size of the item and participant samples, as well as the fact that the similarity in OLD for the similar pairs was less on average here than in the previous experiments.

Plausibly, facilitation of target identification plays a role in the priming effects found for *yes* responses. That is, primes activate the internal memory representations of similar target words and give a head start in target identification. The present results suggest stronger facilitation by beginning rather than end overlap primes. To date we could find no other evidence on this issue in the same-different task. Masked priming in the LDT was examined for these materials in one study by Frisson, Bélanger and Rayner (2014), who found inhibitory priming for end-overlap primes and null effects for beginning-overlap primes, providing some indirect support for stronger facilitation with beginning overlap.

Regardless of the implications for orthographic similarity effects on word identification, the present results provide convergent behavioral validation of the subjective ratings obtained in Experiments 1 and 2. Because primes were masked and unavailable to participants during their decision making, it is unlikely that the results reflect any conscious strategy. Taken together, Experiments 1 to 3 indicate that the similarity structure of English orthography drives both perceived similarity and participants' behavior in the same-different task.

Experiment 4

Experiments 1 to 3 confirmed that OLD does predict similarity ratings, and assessed the role of phonological and beginning overlap in word pairs matched on orthographic similarity as defined by OLD. In Experiment 4, ratings were obtained on a large sample of word pairs, with a view to generalising the previous results to a sample of items that were chosen to cover a large range of similarities. We also examined the relative ability of the OLD metric and two theory-based orthographic similarity metrics to predict ratings. If a metric can be found to account for a large proportion of the variance in mean ratings, then an automated procedure could be devised to estimate the relative similarity of brandnames as perceived by consumers. In addition, the predictive utility of a phonological similarity metric was assessed because orthographic similarity is confounded with phonological similarity. Based on the weak effects of a vowel match, we expected orthographic similarity to be a more important determinant of ratings. The metric used was the Phonological Levenshtein distance (PLD), the phoneme-based equivalent of OLD, which is the only phonological metric available for longer words (Balota et al., 2007).

As before, word pairs were devised to have one target word and a comparison word. The sample of words was chosen to cover a large range of lengths (4 to 12 letters) and with respect to the targets, to approximate their relative frequency in written language. The OLD metric was used to provide a large range of similarities at each target length. For each length there were pairs of maximum similarity (one letter different, respecting position) and maximum dissimilarity (no letters shared in any position), and a range of similarities in between. We assumed that pairs whose members were highly discrepant in length would not typically be considered at risk of confusion; for this reason the majority of targets differed in length from their comparison words by 2 or fewer letters. The distribution of word length and word frequency over OLD values was constrained by the item selection procedures and as a result was not well designed for analyses of the predictive effects of length and frequency. In any case, analyses showed small and mainly nonsignificant effects of these variables; consequently, they are not discussed in Experiments 4 and 5.

Method

Participants. One hundred university students participated for course credit in an introductory Psychology course or for a payment of $10.

Materials and Design. The target words had a mean frequency of 9 per million (range of 1.1 to 39) in the British National Corpus (BNC, Kilgarriff, 1995). The targets (N = 1052) were selected to cover a range of length values in proportions approximating those in the language (as reflected in the BNC). The word length ranged from 4 to 12 letters; the numbers at each length are shown in Table 2. Each target was presented with only one comparison word. These comparison words were chosen to represent maximum, moderate, and very low similarity values as indexed by OLD. They had a large frequency range (0 – 854) with a mean of 28 per million in the BNC and covered a range of lengths relative to the target. The maximum length discrepancies were 6 letters shorter and 4 letters longer than the target, but most pairs (97%) had a length discrepancy of 2 or fewer letters. There were similar proportions of pairs at each OLD value within each target length category, but there was some variation (range of 37 to 52 pairs per category) for 10 – 12 letter words because of errors in calculations and difficulty filling the maximally similar and dissimilar cells. Table 2 shows the number of OLD categories as a function of target length.

Five groups of 20 students rated the orthographic similarity of a subset (N = 263) of the pairs. On average each group had 210 pairs that were unique to the group. The remaining pairs came from a set of 255 pairs were distributed over the groups for a second rating. (A third rating was collected in error for four pairs and was dropped from the analyses.)

Insert Table 2 about here

Procedure. The experiment was conducted as in Experiment 1, except that the rating scale had 6 points (1: dissimilar to 6: most similar).

Results and Discussion

A preliminary check was conducted on rating agreement over the pairs of subgroups for the double-rated items. The absolute value of the discrepancy between pairs of subgroup mean ratings was calculated. The mean discrepancy was .44, which differed significantly from zero, $t(254) = 21.28$, $p < .001$, reflecting that fact that the ratings came from eight different pairings of subgroups of the 100 participants. Despite this difference in mean ratings, the agreement about the relative similarity of the 255 word pairs was high, with the correlation between the two mean ratings at $r = .90$, $p < .001$.

Preliminary inspection of ratings as a function of OLD (unscaled) at each length showed effects of both OLD and length. A one-way ANOVA on ratings for OLD-1 (one-letter-different) confirmed higher ratings as length increased (means of 4.29 for short words and 4.70 for long words), $F(1, 297) = 22.69$, $p < .001$, $\eta_P^2 = .19$. For one-letter different pairs, as target length increases, so too does the proportion of letters shared by the pair members. When length is accounted for by the OLDscaled measure, the one-letter different pairs show decreasing dissimilarity going from short words (.21) to the longest words (.09).

Word beginning- and end-overlap. We assessed the replicability of the Experiment 2 finding of increased similarity ratings when words shared their initial three letters. Word pairs were classified as sharing (at least) the first three letters (N = 143), the last three letters (N = 377), both of these (N = 69), or neither (N = 463). There was no significant difference in OLDscaled for the beginning-overlap (.26) vs. end-overlap (.27) pairs, $F < 1$. The mean number of letters in the overlap was 4.66. Replicating Experiments 1 and 2 over a large range of word lengths and overlap sizes, a one-way ANOVA by items showed that beginning-overlap pairs were rated as more similar than end-overlap pairs, with means of 3.98 and 3.74 respectively, $Fi(1, 517) = 9.12$, $\eta_P^2 = .02$. The advantage for beginning pairs was significant despite the fact that the overlap size was larger for end- than beginning-overlaps, at 4.03 vs. 5.29 letters respectively, $Fi(1, 517) = 49.63$, , $\eta_P^2 = .09$.

An important question that was not addressed in previous experiments is whether end-overlap confers an advantage compared with pairs matching on neither beginning nor end, when OLDscaled is controlled. To address this question we constrained the range of OLDscaled to between 0.15 and 0.72 to make the ranges similar for subsets of end-overlap (N = 302) and no-overlap (neither beginning nor end overlap, N = 219) pairs. These values were chosen because they largely controlled OLDscaled differences without a substantial loss of data. The residual difference in OLDscaled means (.11) was controlled by entering OLDscaled as a covariate in a one-way ANOVA by items. The mean covariate-adjusted similarity rating was higher for end-overlap pairs (3.25) than no-overlap pairs (3.06), $Fi(1, 517) = 9.51$, $\eta_P^2 = .02$. The difference remained significant when pairs with an end-overlap greater than 5 letters were removed, $Fi(1, 410) = 12.58$, $\eta_P^2 = .03$.

Predictors of similarity ratings. The Phonological Levenshtein Distance was converted to a measure between 0 and 1 (PLDscaled) by dividing the distance by the number of phonemes for the pair member with more phonemes. The predictors for analyses were OLDscaled, PLDscaled, and end-weighted and unweighted match values (Davis, 2007) from the Davis Spatial Coding model (Davis, 2010). For the end-weighted match value the default parameter settings were used for dynamic end-letter marking (*Initial Letter Weight =1, c = 1*).

Pairs differing in length by more than 3 letters (fewer than 1% of pairs) were excluded from analyses. The data did not meet the assumptions of multiple regression analysis because the item subsets were allocated to different participant groups. Furthermore, the similarity metrics were highly inter-correlated ($r > .85$) and vulnerable to multicollinearity effects. Consequently we estimated the contribution of each metric separately in a series of linear mixed effects models (Baayen, Davidson, & Bates, 2008). These models operate on the unaggregated (trial-by-trial) data and allow the concurrent assessment of individual participant and item effects. The random intercepts capture differences in mean ratings within conditions (among participants or items) whereas the random slope (here applicable only to participants, because each item falls into a single condition) captures differences over participants in the magnitude of the predictor effects.

Data were analysed using the *lmer* and *lmerTest* functions in R. An estimate of explained variance, the coefficient of determination (R^2), was calculated for each model using the procedure described by Nakagawa and Schielzeth (2013). For each analysis, a series of models were compared: (1) the fullest model, including random intercepts for both participants and items, and random slopes for participants, (2) a reduced model with random intercept and slopes for participant and no random intercept for item, (3) a reduced model with random intercept for participants and items, but no random slopes for participants, (4) a reduced model with random intercepts for item, but no random effects for participant and (5) a reduced model with random intercepts for participant, but no random effect for item. All models were estimated allowing for heteroscedasticity of participant slopes. Models were compared using likelihood ratio tests.

For each of the metrics (OLDscaled, PLDscaled, unweighted and end-weighted match values for the Spatial Coding model), model comparisons showed that model 1, including random intercepts for both items and participants, as well as random slopes for participants, was the best fit to the data. The strongest relationships were observed for Spatial Coding weighted, $t(174.99) = 31.80$, $p < .001$, $R^2 = .42$, and unweighted metrics, $t(197.48) = 30.42$, $p < .001$, $R^2 = .40$, followed by OLDscaled, $t(161.98) = -32.41$, $p < .001$, $R^2 = .36$) and lastly PLDscaled, t(207.11) = -31.5, $p < .001$, $R^2 = .32$. Inspection of the item means (see Figure 4) suggested some possible non-linear effect for Spatial Coding weighted. However the inclusion of these components increased the variance explained by the Spatial Coding weighted score by only a small amount (R^2 increased from .42 to .43).

Insert Figure 4 about here

Experiment 4 confirmed the enhancement of perceived similarity conferred by a beginning-overlap relative to an end-overlap, and further indicated that an end-overlap increases perceived similarity relative to pairs having neither kind of overlap. Additionally, the robust differences in similarity ratings as a function of differences in OLD were replicated.

Participants showed substantial individual variation in their use of the rating scale. This fact was evident in the mixed effects model outcomes for the metrics, which in all cases produced the best fit when random slopes for participants were retained. It was evident also in the mean discrepancy of .44 given for an item set rated by various pairs of participant subgroups. Nevertheless, the mean ratings for participant subgroups showed excellent agreement on the *relative* similarity of pairs, and the means were strongly predicted by the orthographic metrics. The phonological measure, PLDscaled, was highly associated with the orthographic metrics but a less strong predictor, suggesting that it is primarily orthography that is driving participants' ratings.

The Davis (2010) Spatial Coding orthographic metric was superior to OLDscaled, and within the two Davis match calculators, the end-weighted metric was superior. The finding that increasing the weight given to external letters improves the estimate of mean ratings is consistent with a number of sources of evidence about the relative importance of outer letters (particularly the beginning) in word reading. For example, the beginning letters are most informative about word identity (Shillcock, Ellison, & Monaghan, 2000), reading can be accurate when internal letters of words are transposed (as in the Cambridge email mentioned previously), and in a study involving perceptual degradation of letters, readers had a bias towards the outer letters in the early stages of word reading (Beech & Mayall, 2005).

Experiment 5

The findings of Experiment 4 were extended in Experiment 5 to brandnames in an Australian Trademark Register. The primary aims were to confirm in contemporary Australian brandnames the effect of a beginning match, and the relative importance of orthographic similarity metrics in predicting mean similarity ratings. In addition, we assessed whether the practice of including morphemes in brandnames (e.g., *man*, *out, max)* moderated the similarity effect of beginning overlap.

Method

Participants. Forty-two university students participated for course credit in an introductory Psychology course or for a payment of $10. They were divided into two groups of 21.

Materials and Design. Four hundred and thirty pairs of brandnames were chosen from an extract from the Australian Trademark Register of names of products (goods) provided by IP Australia. Due to an unintended repetition, the final total was 427 pairs. Names composed of single letter-strings within the length range 4 to 11 letters were eligible for inclusion in the study (N = 17034 unique names). Only a small percentage (2%) of single-letter-string names fell outside this length range. One pair member was designated the target and the other the comparison word (but for participants this distinction was not evident). The comparison words for each pair were unique, whereas a subset of 48 targets was selected to appear in two pairings, once with a similar and once with an unrelated comparison word. With respect to the lengths of the target and comparison names, 64% of the 807 unique names fell in the length range 6 to 8 letters, 17% in the range 4 to 5 letters, and 19% in the range 9 to 11 letters. As in Experiment 4, most pair members (96%) were 0 to 2 letters different in length, and the maximum length difference was six letters.

The two subgroups rated 247 word pairs each, thus 67 pairs were common to the two lists.

Forty common pairs were the unrelated controls for a comparison of pairs that varied in the beginning vs. end overlap for their similar condition. Similar pairs, which were distributed evenly within each similarity category over the two lists, shared their first three letters (N = 60 pairs) or their last three letters (N = 60 pairs). Within the beginning- and end-overlap pairs, 20 pairs additionally had a morphemic overlap with their comparison name (e.g., OUTBACK–OUTSPAN, AQUAMAX-PROMAX, for beginning- vs. end-overlap respectively). The targets were 7 letters long and the comparison names 5 to 9 letters long. Thirty-six of the 40 unrelated pairs were assigned targets that also appeared in a similar pair, in order to enhance the average comparability of targets over conditions. (One of these 36 targets was replaced by a new target in error). The OLD distance (Keller, 2014), expressed as OLDScaled, was .97 for unrelated pairs and .55 for similar pairs, $p < .001$. There was no significant difference between beginning vs. end pairs in OLDscaled and no similarity x pair type interaction. In a pair type (beginning vs. end) x morpheme (present vs. absent) x similarity ANOVA, there was also no main or interactive effect of the morpheme factor, and no 3-way interaction.

Of the remaining (27) pairs rated by all participants, 24 pairs comprised a set of low similarity pairs (mean OLDscaled = .85) selected as likely to be familiar to the participants. Twelve pairs had members from two different product categories (e.g., GUERLAIN-CONVERSE), and another 12 pairs had the same targets with members from a related product category (e.g., GUERLAIN-AVON). The remaining three pairs were similar names from related product categories; they were too few for separate analysis.

The remaining pairs were selected to represent a range of similarities and lengths within each list. Low similarity pairs were found by random pairings of items. High similarity pairs were orthographic neighbors differing in one or two letters respecting position, for example, WOMBAT-COMBAT; GENERAL-GENERON. Pairs of intermediate similarity were names that shared some of a target word's letters, regardless of position. Target lengths were 4, 5, 6, 8 and 9 letters, and their comparisons were 0, 1, or 2 letters longer. The items were distributed over two lists approximately comparable in length and the scaled OLD distances.

For all 427 pairs, the mean OLDScaled value was 0.62, range 0.1 to 1.0, and the standard deviation was 0.27. The distribution of OLDscaled was different for Experiment 5 compared with Experiment 4, with only 20% of brandname pairs having a scaled distance of less than .3, compared with 39% of the word pairs in Experiment 4. There were correspondingly more moderately dissimilar pairs among the brandnames. This difference resulted largely from the scarcity of highly similar pairs in the brandname database, as well as the different methods used to generate pairs for words and brandnames.

Procedure. The experiment was conducted as in Experiment 4, with a rating scale having 6 points (1: dissimilar to 6: most similar). A rest break was given after every 50 trials.

Results and Discussion

Analyses. Separate analyses were conducted by participants to assess the effects of beginning and end overlap and the product category effect for familiar names. Then all 427 pairs and all participant data were used in analyses to predict participant ratings from the orthographic similarity metrics employed in Experiment 4. Given the absence of pronunciation information, phonological similarity was not included. There are no frequency counts available for the names.

Beginning and end overlap. A Morphemic match (morpheme match vs. not) x Pair type (beginning vs. end overlap) x Similarity (overlap vs. unrelated) ANOVA was conducted on participants' ratings. The mean ratings are shown in Figure 5. There were main effects of Pair type, with beginning pairs rated as more similar overall, $F(1, 41) = 30.55$, $\eta_P^2 = .43$, and Similarity, with similar pairs rated higher than unrelated controls, $F(1, 41) = 224.35$, $\eta_P^2 = .85$. There was a significant Pair type x Similarity interaction, confirming that the difference in ratings between similar pairs and their controls (the similarity effect) was larger for the beginning-match pairs (1.81) than the end-match pairs (1.05), $F(1, 41) = 36.52$, $\eta_P^2 = .47$. There was no main effect of the morpheme variable ($F < 1$), and no interaction of morpheme match with pair type ($F < 1$), and no 3-way interaction ($F < 1$). There was a two-way interaction of Morpheme match x Similarity, $F(1, 41) = 8.55$, $\eta_P^2 = .17$. This result reflects a larger similarity effect for morphemic-match pairs (1.54) than the pairs without a morpheme match (1.33). This interaction was not significant (*p* = .2) in the items analysis, in which all factors were varied between items.

Insert Figure 5 about here

Familiar Brandnames and Product categories. A one-way ANOVA compared the orthographically dissimilar pairs as a function of whether they came from the same/related or different product category. This variable had no effect on the ratings, with means of 1.4 and 1.5 in order for same vs. dissimilar categories, $p = .24$.

Predictors of similarity ratings. The predictors were OLDscaled and end-weighted and unweighted match values from the Davis Spatial Coding model (Davis, 2007). Pairs differing by more than 3 letters in length (1.6% of pairs) were excluded from analyses. As in Experiment 4, we estimated the contribution of each metric and the length variable set separately in a series of linear mixed effects models. All models were estimated allowing for heteroscedasticity of participant slopes; however this model did not converge for OLDscaled and this variable was modelled assuming homoscedasticity.

For each of the distance metrics, model comparisons again showed that model 1, including random intercepts for both items and participants, as well as random slopes for participants, was the best fit to the data. The strongest relationship was observed for the Spatial Coding end-weighted scores, $t(65.38) = 20.45$, $p < .001$, $R^2 = .41$, followed by the Spatial Coding unweighted scores, $t(74.43) = 19.67$, $p < .001$, $R^2 = .38$, and then OLDscaled $t(66.77) = -22.88$, $p < .001$, $R^2 = .35$. Figure 6 shows the scatter plot of mean pair ratings on the Spatial Coding match value.

Insert Figure 6 about here

In summary, Experiment 5 successfully generalised the principal results of Experiment 4 to brandnames. The results strongly confirmed the conclusions of Experiment 4: Overlap in name beginnings enhanced perceived similarity, and although individuals differed in their use of the rating scale, the variation in their mean ratings over item pairs was strongly predicted by the orthographic similarity metrics. The end-weighted match value from the Spatial Coding model (Davis, 2010) again was the best predictor of similarity ratings. The agreement between the two experiments is compelling given the rather different distribution of OLD similarities as a function of length in the word and brandname item sets.

The primary new information provided by Experiment 5 was that perceived similarity in the item set incorporating a beginning or end overlap was incremented by a small amount if similar pairs shared a morpheme. Because this effect was not significant in the items analysis (which has low power because all factors were varied between-items) replication is required. The apparent impact of the morphemes might reflect the contribution of shared meaning to experienced similarity. Finally, it was found that ratings of familiar dissimilar pairs were not affected by their product category (same vs. different), suggesting that functional aspects of the product do not affect perceived similarity of the names. However a limitation on this result was that the item set was small.

General Discussion

Summary. The present series involved three studies of participants' ratings of word pairs, one study in which participants rated the similarity of pairs of brandnames, and one behavioral study in which masked primes preceded targets in a same-different task. As noted, the results were clear-cut. Although individual participants differed in their use of the rating scales, their mean ratings of both word and brandname pairs robustly tracked orthographic similarity as assessed by objective metrics. The best predictor among the metrics was the end-weighted orthographic match calculator from the Spatial Coding model (Davis, 2010).

A match in the stressed vowel had small and somewhat variable effects on similarity ratings and a non-significant priming benefit in the same-different task. Additionally, a metric for phonological similarity was a less successful predictor of ratings than the orthographic metrics. By contrast, for word pairs that were equally similar in the OLD metric, a beginning overlap increased similarity ratings compared with an end overlap (first vs. last three letters, respectively). The effect of beginning overlap was evident in all 5 studies, with the behavioural study (Experiment 3) showing a significantly larger masked priming benefit for beginning overlap primes than end-overlap primes in the latency to judge whether the target was the same as a prior probe word.

Implications for visual word identification. The finding that similarity ratings were well predicted by metrics that allow some positional uncertainty of shared letters is consistent with current research in reading. Recent evidence favors positional flexibility in letter coding over traditional slot-based letter coding schemes of models of visual word identification (Davis, 2010; Gomez, Ratcliff, & Perea, 2008; Grainger & Van Heuven, 2004; Norris & Kinoshita, 2012).

The results for a beginning overlap converge with research on reading from eye-tracking and behavioural studies (Grainger, Granier, Farioli, Van Assche, & van Heuven, 2006; Rayner et al., 2006). This research, together with the Cambridge email, additionally indicates that the end letters of a word may carry more weight in word reading than the middle letters. Consistent with this possibility, the best predictor among the metrics places a higher weight on both beginning and end letters (Davis, 2010). An analysis of the large item set of Experiment 4 revealed that end-overlap pairs were rated as more similar than pairs without a beginning- or end-overlap when OLDscaled was controlled, and also when pairs with end-overlaps of more than 5 letters were excluded. Thus a tentative conclusion, which accords with the eye-tracking results of Rayner and colleagues (2006), is that end-overlap pairs are less similar than beginning-overlap pairs and more similar than pairs without a beginning or end overlap. This conclusion also accords with a masked priming study in which primes consisted of a words' first three or last three letters (Adelman et al., 2014). The present finding that a vowel match has small effects is consistent with the possibility that the orthography-overlap priming effects observed in the lexical literature are orthographic rather than phonological in origin.

The orthographic similarity metric of the Spatial Coding model of visual word identification (Davis, 2010) addressed findings on orthographic similarity in the lexical processing literature, much of which involves pre-conscious effects (masked priming). The fact that this metric aligns closely with subjective ratings suggests a concordance between similarity as revealed in masked priming tasks and subjective impressions of similarity. Perhaps the implication is that subjective similarity is driven by the tendency of a word representation to be activated by another word sharing letters with it. Regardless of the precise nature of the effect, the present results suggest that the driver of at least some of the orthographic similarity effects on lexical processing is also a driver of subjective impressions of orthographic similarity.

Implications for brandname confusion. The present evidence about what makes words orthographically similar converges with behavioural research in word reading and with the predictions of similarity offered by objective metrics. In addition, because prediction by objective metrics was similar for words and brandnames, research findings with words can be applied to issues in law and marketing concerning brandnames.

The results of the present studies have clear implications for practice in law and marketing. The finding that word-initial overlap increases judged similarity validates a long-standing assumption made by courts and by examiners. Our studies have focused on visual presentation, which trademark decision-makers have recognised is important where products or services are likely to selected by consumers from shelves or other visual presentation – a significant proportion of goods and services sold in self-service stores and online. It remains to be established whether word beginnings are equally important in auditory presentation, which can be more important for goods or services ordered or requested orally. Our findings also provide some support for other common assumptions made by trademark decision-makers: That shared or variant endings can impact on similarity, but do so less consistently than shared or variant beginnings, and that common beginning morphemes give a further boost to perceived similarity.

It is important to recognise that our studies do not directly test the ultimate legal question in trademark and related laws: we tested *perceived similarity*, rather than the ultimate harm the law is seeking to guard against, that is, *confusion.* However, the law assumes that similarity is a cause of consumer confusion and purchasing behaviour, and hence both courts and examiners often use visual similarity as a proxy for confusion or, at the very least, as one of the starting points for their consideration. These results are thus directly relevant to the legal assessment.

More generally, examination of students' ratings yields two preliminary conclusions that could be important to the way that trademark decisions are made both in examination and in the context of disputes. The first is that the judgments of a single individual are not necessarily a good guide to the consensus of the group. There was considerable variability in the way that individuals used the rating scales, as reflected in the facts that a maximum of 42% of the total item and participant variance was accounted for by linear effects of objective metrics, and including slopes for participants improved the model fits. The clear implication is that judgments by a single individual in brandname registration decisions or legal disputes – currently the dominant method of decision-making - cannot be said to reflect the judgments of the community of consumers. The fact that one examiner or a trial judge thinks that two words are similar could be quite unreliable and subject to significant variation. The second, complementary, conclusion is that averaging judgments over even a relatively small group (here, approximately twenty individuals) produced robust and reliable estimates of the relative similarity of word and name pairs. The 255 pairs of items that were rated by two subgroups of twenty students each (Experiment 4) showed a high correlation of subgroups' mean ratings, r = .90, even though the composition of the subgroups varied over pairs, and even though there was a discrepancy in the subgroup means for these pairs. Finally, there is no reason to expect differences between the present university students and other consumer samples, given that these perceptions of relative similarity will be grounded in shared cultural and language experiences. These findings suggest then, that obtaining relatively robust measures of similarity need not involve large and expensive surveys.

We recognise there are some complications in operationalizing these ideas in the context of particular disputes. An implication of the present results is that decisions about the similarity of a pair of names will not be identical each time that the judgments of a group are averaged. Thus a firm, *absolute*, measure of similarity is not achievable, given that individual differences, the nature of the rating scale, and the context provided by other items, will affect judgments. Second, by contrast, reliable and robust decisions about the *relative* similarity of pairs can be made by groups of raters. Thus, provided that suitable benchmarks can be included for comparison, it is possible to obtain useful information from group judgments about whether a pair of names is undesirably similar for consumers. The question of what the appropriate benchmarks would be would require further consideration.

Perhaps our most significant finding, with the most immediate practical uses, is that *metrics,* particularly the end-weighted metric from the Spatial Coding Model (Davis, 2010), can provide excellent predictions of average subjective ratings of the relative similarity of words, including brandnames. Experiment 5 produced some evidence that a shared morpheme produces a small increment in rated similarity, plausibly an effect of shared meaning that is not captured by orthographic metrics. Nevertheless, the metric provides a useful estimate of the relative similarities of name pairs and thus could support assessments for trademark registration in particular. A notable feature of trademark registration is that it often involves exactly the kind of process we have undertaken here: namely, simple comparison of words without considering other factors such as colour, font, packaging, or marketing of products. When a company seeks registration of a word (such as a brandname), examiners consider how similar the word is to other words, in relative isolation. It is here our findings could be most relevant. For example, an examiner considering the registration of a new word trademark could efficiently extract from the Trademark Register a pool of existing registrations which, when compared to the new application, exceed a benchmark similarity value. A metric could also produce useful information for making similarity comparisons. For example, a metric could estimate the distribution of similarities and the average similarity within a product category. While courts have hesitated to allow trademark offices to use simple metrics as *the* (only) *basis* for allowing or rejecting registrations, a metric could provide at least an initial list for consideration against other factors not measured by the metric (such as semantic similarity).

The use of metrics in legal disputes involving goods or services marketed to consumers is more complex, because many more factors (packaging, colour, marketing and retailing strategies) come into play. Nevertheless, contested pairs could have their similarity assessed against a benchmark to provide a more robust similarity judgment than that obtainable from a single judge. Given that courts' trademark decisions are sometimes criticised for their inconsistency (Davison & Horak, 2012), the existence of a tool that is reliable, objective and easy to apply is at least worth considering as one of a range of factors, even if it cannot be determinative. An advocate who could show that their words were no more similar according to the metric than existing marks on the register, or, on the other hand, considerably more similar, might not necessarily win their case given the range of factors relevant to the judgment of similarity, but might at least gain some forensic advantage in a dispute. In conclusion, the present findings have significant practical implications for trademark law as it applies to brandnames.

References

Adelman, J. S., Johnson, R. L., McCormick, S. F., McKague, M., Kinoshita, S., Bowers, J. S., . . . Davis, C. J. (2014). A behavioral database for masked form priming. *Behavior Research Methods, 46*(4), 1052-1067. doi:10.3758/s13428-013-0442-y

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language, 35*(6), 775-800. doi:10.1006/jmla.1996.0040

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. doi:10.1016/j.jml.2007.12.005

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon project. *Behavior Research Methods, 39*(3), 445-459. doi:http://dx.doi.org.ezproxy.library.uq.edu.au/10.3758/BF03193014

Beech, J. R., & Mayall, K. A. (2005). The word shape hypothesis re-examined: Evidence for an external feature advantage in visual word recognition. *Journal of Research in Reading, 28*(3), 302-319. doi:10.1111/j.1467-9817.2005.00271.x

Berent, I., & Perfetti, C. A. (1995). A rose is a REEZ: The two-cycles model of phonology assembly in reading English. *Psychological Review, 102*(1), 146-184. doi:10.1037/0033-295X.102.1.146

Burrell, R. G., & Handler, M. (2016). *Australian Trade Mark law* (2nd ed.). Melbourne: Oxford University Press.

Burt, J. S., & Duncum, S. (2017). Masked form priming is moderated by the size of the letter-order-free orthographic neighbourhood. *Quarterly Journal of Experimental Psychology, 70*(1), 127-141.

Chandon, P., Hutchinson, J. W., Bradlow, E. T., & Young, S. H. (2009). Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing, 73*(6), 1-17.

Davis, C. J. (2007). MatchCalc. Retrieved from http://www.pc.rhul.ac.uk/staff/c.davis/Utilities/MatchCalc/index.htm

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117*(3), 713-758. doi:10.1037/a0019738

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance, 32*(3), 668-687. doi:10.1037/0096-1523.32.3.668

Davison, M. J., & Horak, I. (2012). *Shanahan's Australian law of trade marks and passing off*. Pyrmont, NSW, Australia: Thomson Reuters.

Dinwoodie, G., & Gangjee, D. S. (2015). The image of the consumer in European trade mark law. In D. Leczykiewicz & S. Weatherill (Eds.), *The image(s) of the consumer in EU law*. Oxford: Hart Publishing.

Forster, K. I. (1987). Form-priming with masked primes: The best match hypothesis. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 127-146). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 680-698. doi:10.1037/0278-7393.10.4.680

Frisson, S., Bélanger, N. N., & Rayner, K. (2014). Phonological and orthographic overlap effects in fast and masked priming. *The Quarterly Journal of Experimental Psychology, 67*(9), 1742-1767. doi:10.1080/17470218.2013.869614

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115*(3), 577-600. doi:10.1037/a0012667

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language, 35*(5), 623-647. doi:10.1006/jmla.1996.0033

Grainger, J., Granier, J.-P., Farioli, F., Van Assche, E., & van Heuven, W. J. B. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance, 32*(4), 865-884. doi:10.1037/0096-1523.32.4.865

Grainger, J., & Van Heuven, W. J. B. (2004). Modeling Letter Position Coding in Printed Word Perception. In P. Bonin (Ed.), *Mental lexicon: "Some words to talk about words".* (pp. 1-23). Hauppauge, NY, US: Nova Science Publishers.

Huang, V., Weatherall, K., & Webster, E. (2012). The use of survey evidence in Australian trade mark and passing off cases. In A. Kenyon, N. Wee Loon, & M. Richardson (Eds.), *Law of reputation and brands in the Asia Pacific region* (pp. 181-202): Cambridge University Press.

Keller, T. (2014). Fuzzy string processing using Damerau-Levenshtein distance. Retrieved 2014 http://fuzzy-string.com/

Kilgarriff, A. (1995, November 3, 1998). BNC database and word frequency lists. Retrieved from http://www.kilgarriff.co.uk/bnc-readme.html

Kinoshita, S., & Norris, D. (2009). Transposed-letter priming of prelexical orthographic representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 1-18. doi:10.1037/a0014277

Kinoshita, S., & Norris, D. (2010). Masked priming effect reflects evidence accumulated by the prime. *The Quarterly Journal of Experimental Psychology, 63*(1), 194-204. doi:10.1080/17470210902957174

Lupker, S. J., Nakayama, M., & Perea, M. (2015). Is there phonologically based priming in the same−different task? Evidence from Japanese−English bilinguals. *Journal of Experimental Psychology: Human Perception and Performance, 41*(5), 1281-1299. doi:10.1037/xhp0000087

Lupker, S. J., Perea, M., & Davis, C. J. (2008). Transposed-letter effects: Consonants, vowels and letter frequency. *Language and Cognitive Processes, 23*(1), 93-116.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133-142. doi:DOI: 10.1111/j.2041-210x.2012.00261.x

Nakayama, M., Sears, C. R., & Lupker, S. J. (2010). Testing for lexical competition during reading: Fast priming with orthographic neighbors. *Journal of Experimental Psychology: Human Perception and Performance, 36*(2), 477-492. doi:10.1037/a0016800

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General, 137*(3), 434-455. doi:10.1037/a0012799

Norris, D., & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review, 119*(3), 517-545. doi:10.1037/a0028450

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO! Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51*(2), 231-246. doi:10.1016/j.jml.2004.05.005

Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding Wrods With Jubmled Lettres: There Is a Cost. *Psychological Science, 17*(3), 192-193. doi:10.1111/j.1467-9280.2006.01684.x

Schneider, W., Eschman, A., & Zuccolutto, A. (2002). E-Prime user's guide. Pittsburgh, PA: Psychology Software Tools.

Shillcock, R., Ellison, T. M., & Monaghan, P. (2000). Eye-fixation behavior, lexical storage, and visual word recognition in a split processing model. *Psychological Review, 107*(4), 824-851. doi:10.1037/0033-295X.107.4.824

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 26*(6), 608-631. doi:10.1016/0749-596X(87)90105-7

Treiman, R., Kessler, B., Zevin, J. D., Bick, S., & Davis, M. (2006). Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology, 93*(1), 1-24.

Van der Lans, R., Pieters, R., & Wedel, M. (2008). Research Note - Competitive brand salience. *Marketing Science, 27*(5), 922-931.

Van Horen, F., & Pieters, R. (2012). Consumer evaluation of copycat brands: The effect of imitation type. *International Journal of Marketing, 29*(3), 246-255. doi:10.1016/j.ijresmar.2012.04.001

Walsh, G., Mitchell, V. G., Kilian, T., & Miller, L. (2010). Measuring consumer vulnerability to perceived product-similarity problems and its consequences. *Journal of Marketing Management, 26*(1-2), 146-162.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review, 8*(2), 221-243. doi:10.3758/BF03196158

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*(5), 971-979. doi:10.3758/PBR.15.5.971

Appendix A

Similar pairs for the vowel-overlap set in Experiment 1 (OLD = 2).

Target (Six letters)	Similar word, different vowel	Target (Six letters)	Similar word, same vowel
	Five letters		Five letters
seized	sizes	fallen	false
launch	laugh	driven	risen
device	devil	quoted	voted
poured	pound	attend	trend
barely	badly	smooth	shoot
nation	ratio	handle	angle
horror	error	thirty	shirt
glance	grace	spoken	smoke
talent	alert	muscle	uncle
circle	cycle	thrown	grown
fierce	fence	cheese	cheek
marine	arise	behave	brave
module	mouse	waited	aimed
varied	valid	strain	grain
reveal	rival	switch	pitch
walker	baker	freely	feels
salary	alarm	closer	chose
fought	rough	plenty	penny
gender	genes	liable	bible
parent	agent	stream	treat
	Six letters		Six letters

motion	cotton	bitter	fitted
stolen	styles	agents	agenda
rarely	namely	faster	farmer
sudden	hidden	dealer	deeper
casual	visual	slight	lights
lesser	losses	clever	eleven
retain	repair	clause	causes
retail	recall	stable	tables
settle	cattle	deeply	weekly
wished	sighed	copper	copies
manual	mutual	resist	insist
golden	wooden	stayed	stages
	Seven letters		Seven letters
remove	resolve	gained	trained
inland	islands	honest	contest
mature	mixture	obtain	contain
tested	twisted	shaped	escaped
mostly	monthly	parish	spanish
secure	lecture	resort	restore
finest	fitness	warned	awarded
assess	possess	remote	promote
softly	shortly	silent	violent
orange	arrange	combat	compact
intent	instant	chapel	channel
priest	protest	ticket	cricket
praise	promise	relate	replace

bother	mothers	denied	derived
severe	reverse	holder	soldier
gently	greatly	wealth	healthy
stupid	studied	stored	stories
cousin	causing	export	explore
leaned	cleared	vision	mission
aspect	suspect	harder	charter

Appendix B.

Similar pairs for the beginning- vs. end-overlap set in Experiment 2 (OLD = 2).

Target (Six letters)	Similar word, beginning same	Target (Six letters)	Similar word, end same
thirty	thick	seized	gazed
holder	holes	walker	baker
motion	motor	shaped	wiped
priest	prize	behave	grave
spoken	spoon	remove	prove
switch	swing	slight	ought
waited	waist	quoted	dated
marine	marry	bitter	utter
stupid	stuck	stored	dared
clause	clash	muscle	cycle
stayed	stamp	tested	voted
cousin	count	poured	cared
honest	honey	launch	bench
closer	clock	liable	noble
wished	wisdom	gained	banned
remote	remark	settle	castle
talent	taller	assess	excess
thrown	thrust	obtain	domain
strain	stroke	sudden	wooden
module	modest	clever	server
bother	bottle	handle	needle
resist	rescue	vision	nation

mostly	mosaic	inland	expand
barely	barrel	export	cohort
silent	silver	secure	endure
fought	fourth	golden	burden
stable	stairs	circle	oracle
praise	prayer	intent	urgent
leaned	leather	parish	rubbish
manual	mansion	resort	comfort
gender	genetic	warned	stained
stream	strings	salary	summary
combat	compete	aspect	neglect
driven	drifted	severe	nowhere
attend	attract	deeply	sharply
harder	harvest	parent	comment
faster	fashion	reveal	conceal
plenty	pledged	dealer	simpler
lesser	lessons	softly	firstly
retain	retreat	rarely	vaguely
wealth	weather	varied	studied
horror	horizon	stolen	stomach

Author Note

The research was supported by Linkage Grant LP120100249 from the Australian Research Council.

We thank Philippe Lacherez for assistance with modelling and Walter Leggett for assistance with item preparation.

Table 1

Experiment 3: Mean error rates for the vowel and beginning-end sets in the same-different task, yes responses (same trials).

Prime type	Vowel		Overlap	
	Match	Mismatch	Beginning	End
Similar	2.8	2.8	3.5	2.8
Unrelated	3.5	5.6	3.5	4.9

Table 2

Experiment 4: Characteristics of the target words.

	Length (letters)								
	4	5	6	7	8	9	10	11	12
N	84	120	144	144	140	140	120	92	68
%	8	11	14	14	13	13	11	9	7
No. OLD categories	3	3	4	4	5	5	6	6	6

Figure captions

Figure 1. Experiment 1: Mean similarity ratings (range 1 – 4) as a function of vowel match and target length, for the critical similar pairs vs. unrelated controls. Error bars show the standard error of the mean.

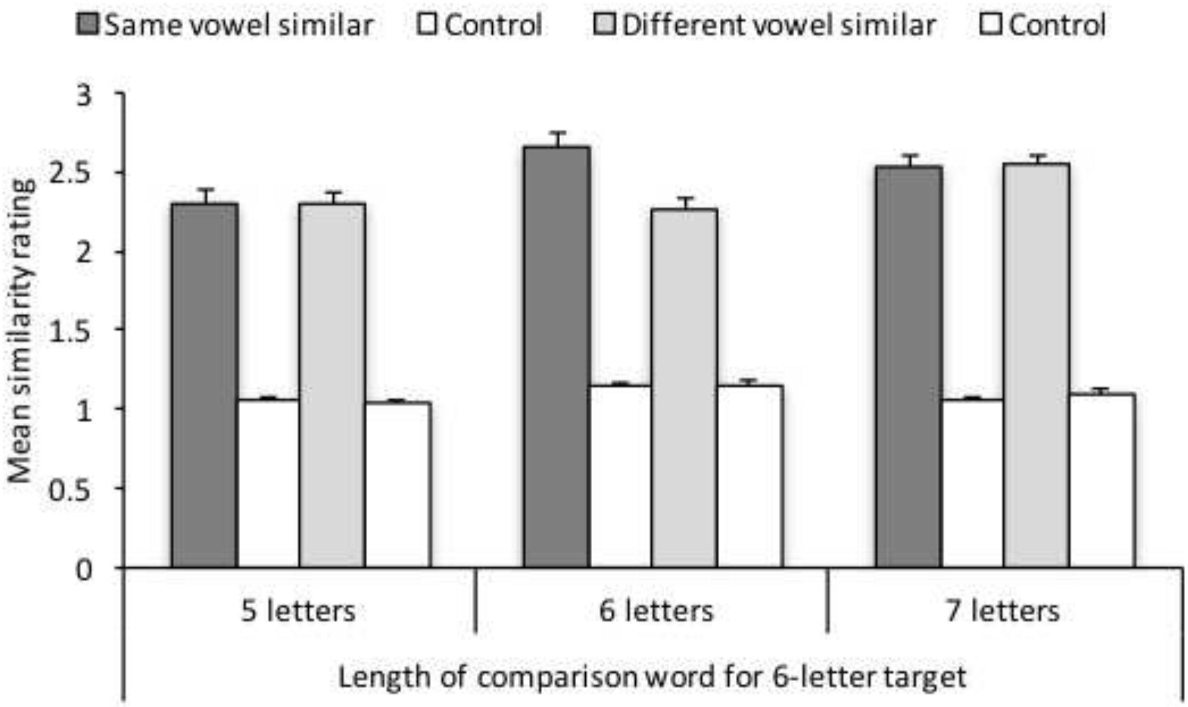
Figure 2. Experiment 2: Mean similarity ratings (range 1 – 4) as a function of beginning vs. end overlap and comparison word length, for the critical similar pairs vs. unrelated controls. Error bars show the standard error of the mean.

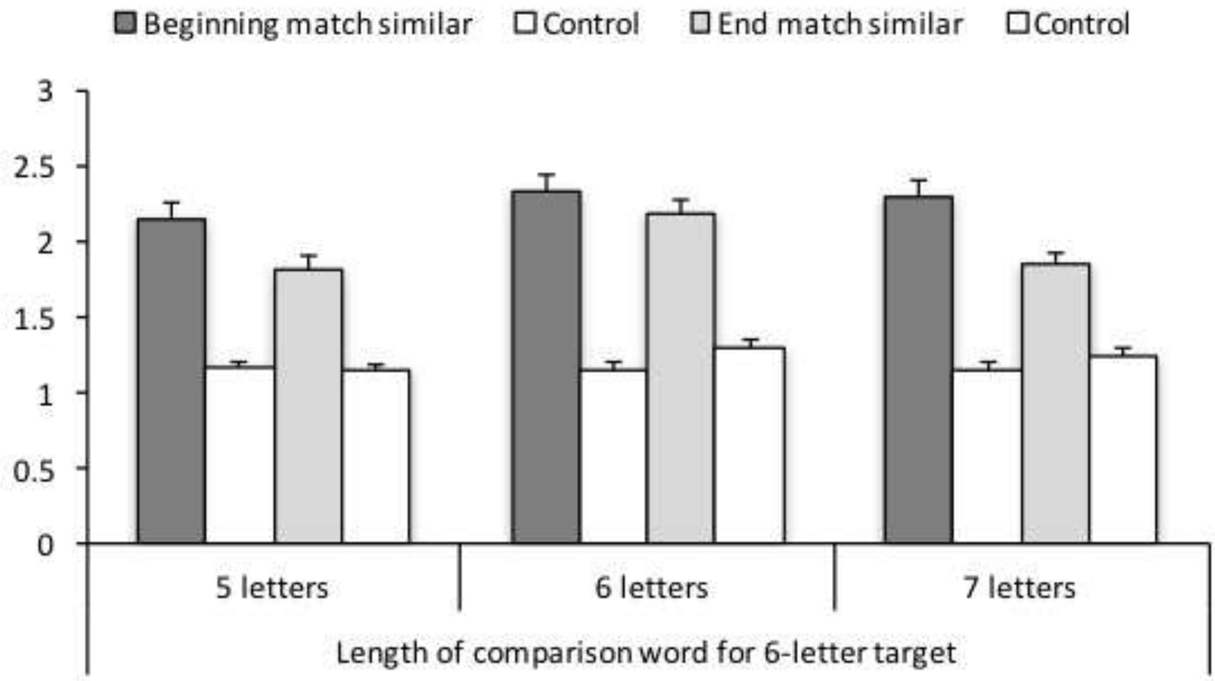
Figure 3. Experiment 3: Mean same response latencies (ms) as a function of beginning vs. end overlap and vowel match vs. mismatch for the similar vs. unrelated prime conditions in the masked priming same-different task. Error bars show the standard error of the mean.

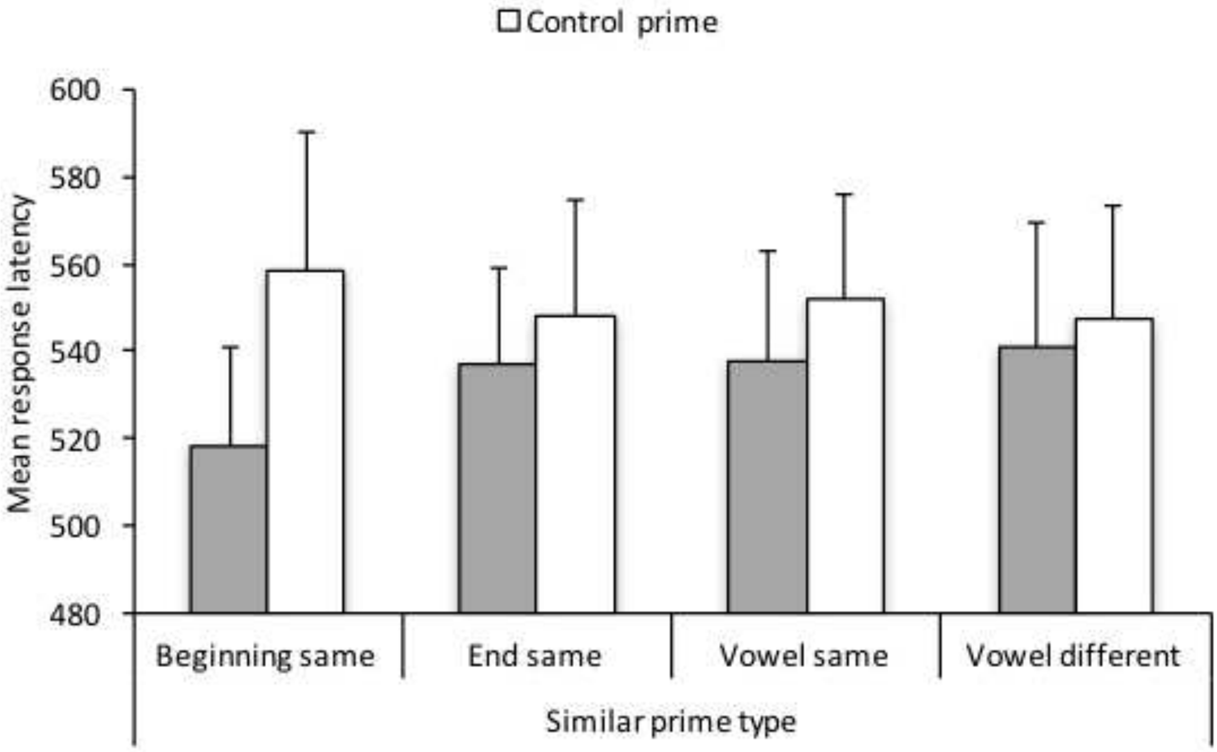
Figure 4. Experiment 4: Scatter plot of mean similarity ratings for the word pairs as a function of their similarity on the Endweighted Match of the Spatial Coding model.

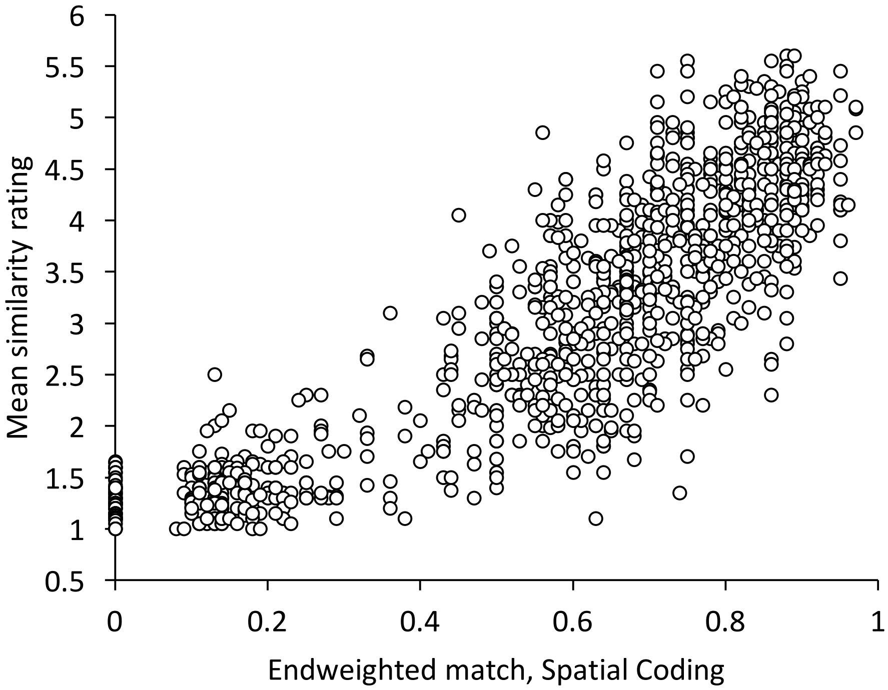
Figure 5. Experiment 5: Mean similarity ratings of brandname pairs (range 1 – 6) as a function of orthographic similarity, beginning vs. end overlap for similar pairs, and whether the overlap was a morpheme vs. not. Error bars show the standard error of the mean.

Figure 6. Experiment 5: Scatter plot of mean similarity ratings for the brandname pairs as a function of their similarity on the Endweighted Match of the Spatial Coding model.

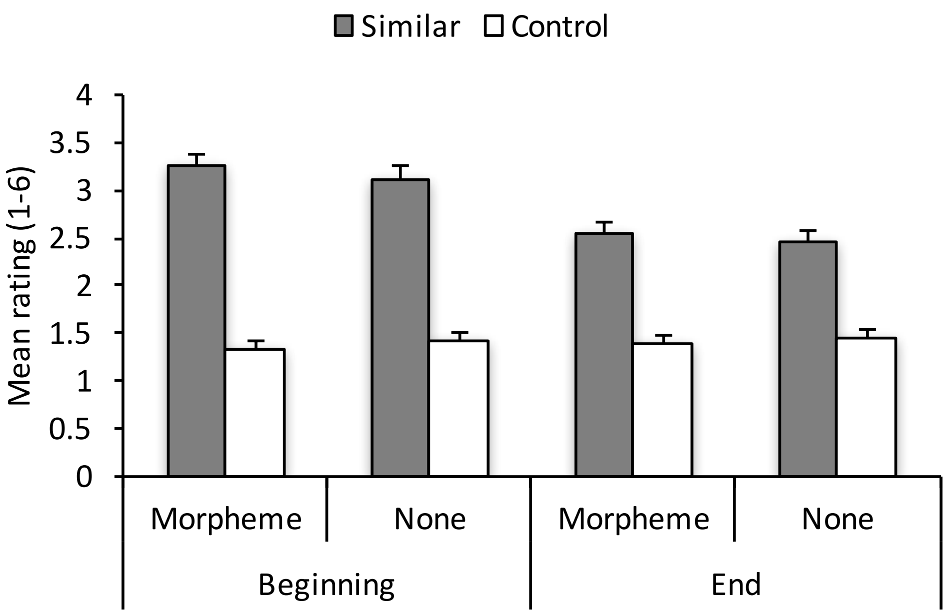

Same vowel similar
Control
Different vowel similar
Control
3
2.5
2
1.5
1
0.5
0
Mean similarity rating
5 letters
6 letters
7 letters
Length of comparison word for 6-letter target


Beginning match similar
Control
End match similar
Control
3
2.5
2
1.5
1
0.5
0
5 letters
6 letters
7 letters
Length of comparison word for 6-letter target

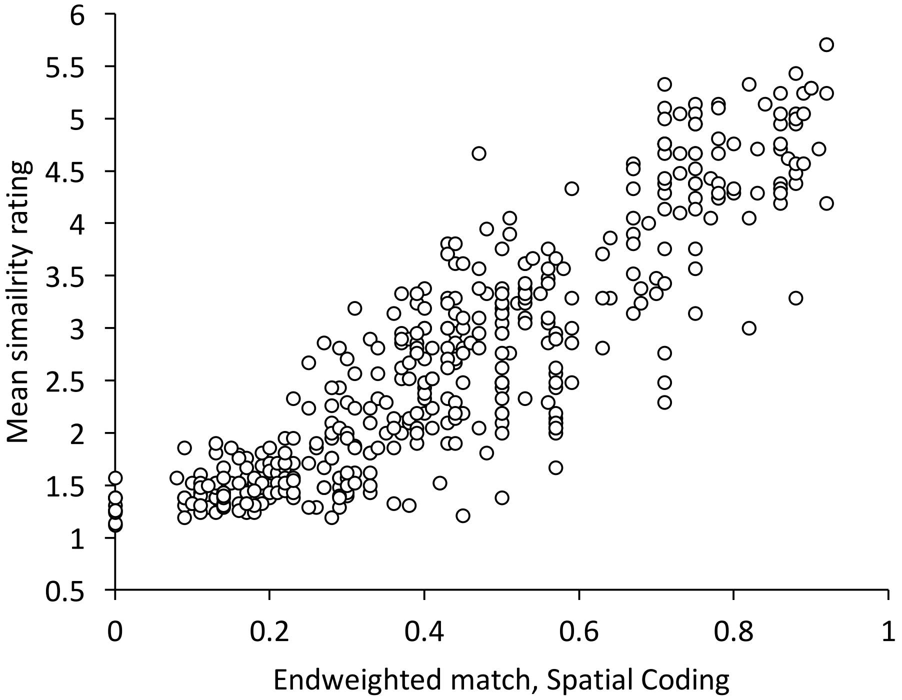

Control prime
600
580
560
540
520
500
480
Mean response latency
Beginning same
End same
Vowel same
Vowel different
Similar prime type



Mean similarity rating
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0.2
0.4
0.6
0.8
1
Endweighted match, Spatial Coding



Similar
Control
Mean rating (1-6)
4
3.5
3
2.5
2
1.5
1
0.5
0
Morpheme
None
Morpheme
None
Beginning
End



Mean simailrity rating
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0.2
0.4
0.6
0.8
1
Endweighted match, Spatial Coding