Article:

Trifonova, Iliyana Venelinova and Adelman, James (2022) Repeated letters increase the ambiguity of strings : Evidence from identification, priming and same-different tasks. Cognitive Psychology. 101445. ISSN 0010-0285

https://doi.org/10.1016/j.cogpsych.2021.101445

Repeated letters increase the ambiguity of strings: Evidence from identification, priming and same-different tasks

Iliyana V. Trifonova[a,b] and James S. Adelman[a]

[a]University of Warwick, [b]University of York

Corresponding author:

Iliyana V. Trifonova

Department of Psychology

University of York

York

YO10 5DD

United Kingdom.

iliyana.trifonova@york.ac.uk

Abstract

Letters are often repeated in words in many languages. The present work explored the mechanisms underlying processing of repeated and unique letters in strings across three experimental paradigms. In a 2AFC perceptual identification task, the insertion but not the deletion of a letter was harder to detect when it was repeated than when it was unique (Exp. 1). In a masked primed same-different task, deletion primes produced the same priming effect regardless of deletion type (repeated, unique; Exp. 2), but insertion primes were more effective when the additional inserted letter created a repetition than when it did not (Exp. 3). In a same-different perceptual identification task, foils created by modifying a repetition, by either repeating the wrong letter or substituting a repeated letter, were harder to reject than foils created by modifying unique letters (Exp. 4). Thus, repetition effects were task-dependent. Since considering representations alone would suggest repetition effects would always occur or never occur, this indicates the importance of modelling task-specific processes. The similarity calculations embedded in the Overlap Model (Gomez et al., 2008) appeared to always predict a repetition effect, but its decision rule for the task of Experiment 1 allowed it to predict the asymmetry between insertions and deletions. In the Letters in Time and Retinotopic Space (LTRS; Adelman, 2011) model, repetition effects arise only from briefly presented stimuli as their perception is incomplete. It was therefore consistent with Experiments 2-4 but required a task-specific response bias to account for the insertion-deletion asymmetry of Experiment 1.

Keywords: visual word recognition; repeated letters; masked priming; orthographic processing; computational modelling; letter processing

1. Introduction

Processing visual information in alphabetic languages requires the successful identification of constituents within multiple hierarchical levels. The identification of a word in a sentence relies upon the recognition of the letters and their order within the word unit. The encoding of letter identities and their position has been the focus of much orthographic processing research. This research has investigated how the visual system encodes letter information and discriminates between several word candidates before selecting (the correct) one from a set of representations. The results of this research effort have led to the development of letter and word processing theories that have been explicitly formulated in the architecture of computational models of visual word recognition tasks. The mechanisms of letter identity and position encoding are implemented in the encoding schemes of these models and determine both the set of considered candidates and the difficulty of the process of selection.

Key findings about how letter information is processed when letters are embedded in strings include: the activation of abstract letter representations is not affected by differences in font, case and size (e.g. Bowers, Vigliocco, & Haan, 1998; Kinoshita & Kaplan, 2008); letter strings are processed in a specialized manner that is different from the one associated with other domains such as symbol strings and there is a processing advantage for initial positions (Aschenbrenner, Balota, Weigand, Scaltritti & Besner, 2017; Jordan, Thomas, Patching, & Scott-Brown, 2003; Scaltritti & Balota, 2013; Scaltritti, Dufau, & Grainger, 2018; Schoonbaert & Grainger, 2004; Tydgat & Grainger, 2009); and the encoding of letter information is initially imprecise and there is a high degree of positional uncertainty in early word recognition stages (e.g. Kinoshita & Norris, 2009; Lupker, Perea, & Davis, 2008; Peressotti & Grainger, 1999; Schoonbaert & Grainger, 2004; Van Assche & Grainger, 2006; Welvaert, Farioli, & Grainger, 2008). Less clear is how letter information is processed when the same letter is present more than once in one string; evidence has been mixed as to the existence of effects of repeated letters in word identification. Such cases are however important because words with letter repetitions are more common than those without in many languages including English, French and Dutch (Trifonova & Adelman, 2019).

Recently, Trifonova and Adelman (2019) have provided evidence that the presence of repeated letters in words delays their visual processing. This was demonstrated with a regression approach on megastudy data of two key visual word recognition tasks: lexical decision and word naming. In these tasks, a letter string is presented on a computer screen and a response time is measured from the onset of the presentation. In lexical decision, participants indicate whether the letter string forms a genuine word or not by pressing one of two corresponding keys, whereas in word naming, they read the word aloud. Trifonova and Adelman found an inhibitory effect of letter repetition in three different languages, English, Dutch and French. The effect was most robust for cases in which the repeated letters appeared in close proximity but not in immediate adjacency.

These findings are reminiscent of several previously reported perceptual phenomena suggesting interference in cases of repeated visual information. These include: the *repeated-letter inferiority effect* (Bjork & Murray, 1977; Egeth & Santee, 1981) describing poor performance in cases of repeated target letters in tachistoscopic letter identification studies; the *homogeneity effect* (Mozer, 1989) demonstrating lower accuracy in identifying the number of target letters in cases of repetitions; the *redundancy masking effect* (Sayim, & Taylor, 2019) describing a similar effect of perceived smaller numbers of letters in peripheral vision; the *repetition blindness effect*, reporting lower accuracy in detecting several instances of the same word or letter type (Kanwisher, 1987; 1991; Luo & Caramazza, 1996). Regardless of the persistent traces of a repeated letter phenomena reported in various cognitive processing studies, previous factorial studies investigating orthographic processing and effects of repeated letters on the processing of whole letter strings have provided mixed evidence. It is, therefore, still unclear whether the repetition of a letter embedded in a letter string could change the processing difficulty and affect reading.

The goal of this work was to provide further evidence required for the better understanding of the processing of repeated letters embedded in letter strings, as they would appear in words. A more detailed exploration with factorial approach would allow a comparison with the evidence for repeated letter effect from the regression approach on large datasets (Trifonova & Adelman, 2019), as well as with previous inconsistent findings from the orthographic processing literature related to the same research question. An overview of these findings will be shortly presented. This will be followed by a summary of the theoretical assumptions and modelling of repeated letters in visual word recognition. The second section will present a more detailed description of two computational models of visual word recognition: the Letters in Time and Retinotopic Space model (LTRS: Adelman, 2011) and the Overlap Model (Gomez, Ratcliff, & Perea, 2008) and their assumptions regarding the processing of repeated letters. This will be followed by the presentation of four experiments employing several orthographic processing methodologies.

The results of the experiments will be accompanied by predictions from LTRS and the Overlap Model and discussed in the context of both frameworks.

Previous findings

Unlike Trifonova and Adelman's (2019) regression approach, it has been more common to approach the investigation of letter position and identity encoding in strings using factorial experiments. In this approach, experiments are designed with a controlled selection of stimuli, and the effect of interest is explored with experimental manipulations and comparison between contrasting conditions. The experimental paradigms in visual word recognition studies, especially those focused on early perceptual processing, often involve the presentation of some brief perceptual event, such as the target letter string in perceptual identification tasks (e.g., Gomez et al., 2008), or a prime in a masked-primed lexical decision task (Forster, Davis, Schoknecht, & Carter, 1987). In the masked-priming paradigm, participants are typically unaware of the presence of the prime that is briefly presented after a mask (######) and prior to the target. Nevertheless, response times are often faster when the prime is related in form to the target (e.g., desihn-DESIGN) in comparison to when it is not (voctal-DESIGN) (e.g., Adelman et al., 2014). The difference between the two conditions, the priming effect, is usually interpreted as reflecting the degree to which the orthographic codes between the related prime and the target overlap. That is, higher similarities between prime and target are expected to produce more (facilitatory) priming.

Masked-priming lexical decision studies investigating repeated letters in the past have taken advantage of previous findings that primes formed by disrupting the precise letter order, for example by deleting letters from the target base word (relative position subset or deletion primes, e.g., blcn-BALCON), still produce form priming effects (e.g., Peressotti & Grainger, 1999), contrary to the predictions of a strict position-specific encoding scheme (McClelland & Rumelhart, 1981). Primes formed by deleting a repeated or a unique letter from a French base word (e.g., balace vs. balnce for target BALANCE) produced the same form priming effect as each other when compared with a control unrelated six-letter word-like prime (e.g., fodiru-BALANCE; Schoonbaert & Grainger, 2004). These results were, however, inconsistent with the main effect of target type reported in the same study: Targets with repeated letters were more difficult to process. In a different series of experiments, again no effects of repeated letters were reported with primes in which one or two additional repeated or unique letters were inserted (insertion primes: e.g., jusstice, juastice, jussstice, jurqstice - JUSTICE), rather than deleted from the target (van Assche & Grainger, 2006). These insertion primes produced the same priming effect as each other, relative to unrelated primes and did not differ significantly from identity primes (justice). However, the fact that no priming cost was reported with the insertion of one or two letters in the prime suggests that the experiments of van Assche and Grainger (2006) could have been underpowered. Evidence for larger cost with increased number of letter insertions was later provided by Welvaert, Farioli, and Grainger (2008). Furthermore, Adelman et al. (2014) showed that primes with one redundant doubled identity (e.g., deshhign - DESIGN) were less disruptive to the target recognition than two redundant letter identities difference (e.g., desaxign - DESIGN). The results with the masked-priming lexical decision task are therefore inconclusive as to whether repeated letter effects could be observed, and whether such effects could depend on factors such as adjacency or number of repeated letters.

Indeed, Norris, Kinoshita, and van Casteren (2010) argued that repeated letter effects should be more robustly observed when the repeated letters are adjacent. They also suggested that the masked-priming lexical decision task might not be appropriate for the investigation of these effects and argued that the (masked-priming) same-different task might be a more appropriate tool for small sublexical effects, as it is insensitive to lexical factors and the priming effect in this task is suggested to reflect only orthographic processes (Kinoshita & Norris, 2009; Norris & Kinoshita, 2008; Norris et al., 2010). The latter argument, however, was recently challenged by evidence in the literature suggesting a partial phonological contribution (Lupker, Nakayama, & Perea, 2015; Lupker, Nakayama, & Yoshihara, 2018). Nevertheless, priming effects are often larger in the same-different task than the lexical decisions task, so there may be more power in such studies. In the same-different task, a lowercase letter string serves as a temporary reference presented for about 1s simultaneously and above a mask of symbols (#######). This event is succeeded by the presentation of a prime that appears at the place of the mask (in lowercase), followed by the target (in uppercase). The task of the participants is to indicate whether the reference and the target are the same or not, disregarding their case difference. Norris et al. (2010) provided evidence that repeated letter effects could be observed with the masked priming methodology, particularly when combined with the same-different task. They showed greater priming by primes such as *uueer* compared to *ulger* for the target UNDER, as well as by *anex* for ANNEX compared to eupt for ERUPT. They took this evidence as a support of a mechanism of “leakage” of a letter identity to nearby positions due to early perceptual uncertainty. However, the authors did not argue that such effects could be generalized to nonadjacent repetitions.

Evidence with factorial designs for the presence of repeated letter effects in cases of nonadjacent repetitions as well as adjacent repetitions, however, has been provided by Gomez et al. (2008), who used a two-alternative forced choice perceptual identification task. Participants had to identify a briefly presented five-letter nonword target by choosing between the correct option and a foil option. Higher similarities between target and foil were expected to produce lower accuracy. Accuracy was extremely low when the target contained letter repetition in both adjacent and nonadjacent cases. Accuracy was significantly higher when the foils and not the target contained a letter repetition and intermediate when both the target and the foils had a repeated letter. The authors speculated that participants had a bias towards choosing a string without repetition and suggested that the repeated letter effect might be due to repeated letters being difficult to detect at early perceptual stages. However, Gomez et al. did not proceed to examine this effect in more detail.

Taken together, masked-priming paradigm studies in the past have provided some evidence for differential processing between unique and repeated letters for adjacent repetitions (Adelman et al., 2014; Norris et al., 2010), but not for nonadjacent cases (Schoonbaert & Grainger, 2004; van Assche & Grainger, 2006). In addition, the evidence also raises methodological concerns regarding the sensitivity of the masked-primed lexical decision task to effects operating on a letter level, such as the repeated letter effects and an advantage of the masked-primed same-different task for detecting such processes. Repeated letter effects in cases of nonadjacent repetitions have been observed on word naming and lexical decision response times with regression approach on megastudy data (Trifonova & Adelman, 2019) and with a factorial approach in a lexical decision (Schoonbaert & Grainger, 2004) and perceptual identification experiments (Gomez et al., 2008). Further evidence in this area seems to be required to resolve the inconsistent findings and to establish how repeated letters in nonadjacent positions affect the processing of the whole string unit, how their processing differs from that of corresponding unique letters.

The present experiments

The present work investigated the effects of repeated letters in more detail with a factorial approach. We aimed to further understand how the previous inconsistencies in repetition effects might arise from differences in task demands, task sensitivity, and orthographic context. We focused on the under-studied non-adjacent repetitions, using exclusively nonword stimuli to limit the influence of lexical effects. Across our four experiments, we used three tasks believed to be sensitive to early perceptual processing and enhanced our power with larger than usual participants numbers. To interpret the results and compare models, we also modeled these stimulus relationships and (where possible) particular tasks with several models of visual word recognition.

The first of our four experiments used the two-alternative forced choice perceptual identification task. This experiment focused on testing how the visual system keeps track of the number of occurrences of letter identities. Experiments 2 and 3 employed the masked primed same-different task as a comparator for Experiment 1 with a task less susceptible to response biases. The first three experiments included conditions with asymmetric orthographic relationships, in which one of the stimuli has either a deleted or an additional inserted letter, and the deletion or insertion may be a unique or repeated letter. Experiment 4 continued the investigation of the effect with new conditions in which the orthographic pair comprises of stimuli of the same length, and the manipulations include letter replacements. The task in Experiment 4 was a hybrid between the previous two paradigms, a same-different perceptual identification task (without priming). This further explored the processing of strings with repeated letters and tests whether the results we obtain generalize to different orthographic contexts and another paradigm.

Previous modelling

Theories of visual word recognition, like the evidence in the literature, have been inconsistent in the way they treat repeated letters. An example of such inconsistency comes from successors of the Interactive Activation (IA) Model (McClelland & Rumelhart, 1981) such as open bigram models (Grainger, Granier, Farioli, van Assche, & van Heuven, 2006; Grainger & van Heuven, 2003) and the Spatial Coding Model (SCM; Davis, 2010) and its predecessor SOLAR (Davis, 1999). These models were designed to overcome the limitations of the absolute position encoding scheme of their predecessor and to explain masked-priming data reporting strong priming effects from form-related primes with different cases of absolute letter order violations, and they have not been applied to tasks in which a briefly presented string is to be identified[1].

[1] Moreover, although the McClelland and Rumelhart (1981) model was often applied to the identification of *letters* in nonwords, and a mechanism was described for deciding between two different word alternatives, these mechanisms do not readily generalize to decisions between two

The commonality between these interactive activation models is the assumption that the process of selecting the right candidate for among words in the vocabulary (lexical selection) is mediated by mechanisms (implemented with the same equations as McClelland and Rumelhart) in which input information leads to activation of consistent candidates which further compete for selection, while inconsistent candidates are inhibited early on. The consistent candidates that are activated may include some that are only partially consistent with the input information, and they are only partially activated. These models therefore assume that the matching between an input string and the candidates is graded, and the match is affected by the amount of both the consistent and inconsistent information. One of the major differences between these models is their encoding schemes.

With respect to detecting letter repetitions, the Interactive Activation Model (McClelland & Rumelhart, 1981) is insensitive, as the letter identity information operates independently for each letter position. The open bigram model (Grainger & van Heuven, 2003) is sensitive to repetitions. This model encodes letter position and identity through an intermediate set of repesentations (open bigrams) formed by two-letter cobinations with a correct relative left-to-right order. As the word representations are represented by the unique set of open bigrams and duplications of letters lead to duplications of open bigrams, letter repetitions can affect the predictions of the model via the calculation of the matching between the input and the candidates. The Spatial Coding Model (Davis, 2010) uses letter representations for identities and letter positions are represented by spatial patterns. The same letter representations can be used for the same letter in different positions. However, the model has an explicit mechanism that deals with letter repetitions to ensure the model effectively treats repeated letters in the same way as different ones. Therefore, this model is not sensitive to letter repetitions (see Trifonova & Adelman, 2019, for a more detailed discussion of these models and their relevance to repeated letter effects). It is unclear how these models should be extended to deal with identification of briefly presented nonwords or priming to nonwords (in tasks other than lexical decision) given the centrality of the lexical matching process, but it is possible to examine the numerical match that would drive activation for analogous situations in the priming of words. Such “match scores” are commonly examined to understand the predictions of these types of models, although the exact predictions can be influenced by other aspects of the aforementioned lexical matching process.

---

nonword alternatives, because (like the successor models) decisions are based on activity in the permanent lexical representations.

Another model designed to overcome the limitations of the position specific encoding scheme is the Overlap Model (Gomez et al., 2008). This model implements positional uncertainty by representing letter identities in a brief letter string as normal distributions along a letter position dimension. The distributions are centered on the letter positions and their standard deviations are free parameters with the initial position typically having a smaller standard deviation (less noise) than the other positions. The model asseses the similarity between two letter strings by summating the areas of overlap under the curves of the brief string that match a comparison string whose representation is uniformly spread within the relevant position window (i.e., ±0.5 from the center). The two-alternative forced-choice perceptual identification task is implemented by first applying a power function to the overlap scores so-calculated, and then applying the Luce (1959) choice rule to the resulting values. Gomez et al. demonstrated that the model could account for effects such as letter transpositions, insertions, deletions and indeed repetitions, as in these cases a consistent letter identity spills over, or *leaks* to nearby positions and still contributes for a higher overlap score relative to inconsistent letter identity, despite its incorrect position. Gomez et al. also (in an appendix) ordinally compared overlap scores from the model to priming effects from a handful of studies, much as match scores from interactive activation based models as have been, but these investigations did not extend to studies of letter repetitions.

An alternative theory of orthographic processing that is also not based on the IA framework and that has predicted visual processing differences between repeated and unique letters was proposed by Adelman (2011). In his model, Letters in Time and Retinotopic Space (LTRS), the process of word identification and both the observed masked priming and perceptual (tachistoscopic) identification effects are explained by the means of gradual accumulation and availability of visual information in time. Unlike the IA models, priming effects do not result from a graded match that weighs the consistent and inconsistent information and activates the target to a proportionate degree. Rather, in LTRS, inconsistent information is treated as clear evidence that a stimulus is not a particular target.

According to LTRS (Adelman, 2011), during perceptual processing of the prime, several potential targets can be temporarily consistent with the prime, which causes lexical processing of those targets. When any inconsistency is detected between the prime and a potential target, that target no longer undergoes further lexical processing. Perceptual processing of the prime continues (until the prime offset) so that any remaining potential targets can be evaluated. It is the timing of the detection of inconsistency (or timing of the detection of target onset if no inconsistency is detected) that determines the size of the priming effect that target would receive. In LTRS this timing is probabilistic, not deterministic.

Bigger discrepancies from the target give more opportunities for earlier detection of inconsistency and therefore lead to smaller priming effects. Differences in cases of repeated letters are predicted in some conditions: While a unique excess letter is immediately discrepant whenever it is perceived, a repeated letter cannot be perceived to be in excess – and hence discrepant – without further information, either about its twin, or (for non-adjacent repetitions) its local context of other letters. Adelman (2011) showed that LTRS accounts for the repeated-letter results in tachistoscopic identification of Gomez et al. (2008), and in primed lexical decision of Schoonbaert and Grainger (2004) and Norris et al. (2010) and argued that for Van Assche and Grainger's (2006) primed lexical decision experiments, the observed null results were within the predicted range, given the power and precision of those experiments and the small average effects predicted by LTRS for these conditions.

Present modelling

The data from the present experiments will be fitted with LTRS (Adelman, 2011), and the Overlap Model (Gomez et al., 2008). The models will be evaluated across several paradigms – two-alternative forced choice perceptual identification, masked priming (in a same-different task), and perceptual identification (again in a same-different task). Additional implementations of the models for these tasks will be presented where necessary. In addition to the simulations with LTRS and the Overlap Model, successors of the Interactive Activation model (McClelland & Rumelhart, 1981) will also be evaluated with match scores. The match scores calculations will be obtained with Davis's Match Score calculator (http://www.pc.rhul.ac.uk/staff/c.davis/utilities/matchcalc/index.htm) for the SOLAR model (Davis, 1999), as well as for two versions of the overlap bigram scheme: open bigram (Grainger & van Heuven, 2003), and the overlap open bigram (Grainger et al., 2006). The predictions of all models will be presented along with the results of each experiment. In addition, those predictions and the key findings will be summarized in a table (Table 14) and will be further discussed in the General Discussion section.

In the next section, we will present a more formal description of LTRS (Adelman, 2011) and the Overlap Model (Gomez et al., 2008) and their application to the tasks used in this work. This will be followed by the presentation of the four new experiments.

2. Models

2.1. LTRS

In the below, we explain the Letters in Time and Retinotopic Space model (LTRS: Adelman, 2011) explicating several aspects that were implicit in the original paper, including explicit formulae for the derivation of numerical predictions and demonstration of the use of explicit listing of possible perceptual states to derive qualitative parameter-free predictions (in those cases where the model makes such predictions).

LTRS is a model in which different elements of perceptual information arrive at random times to form discrete perceptual states. Each stimulus can lead to some perceptual states but not others. Elements of information about different letters arrive independently, so the sequence of perceptual states is not entirely predictable, and different states are passed through on different presentations of the same stimulus. The various intermediate states in which not all perceptual information has arrived are ambiguous between several possible stimuli.

When a stimulus presentation is brief and post-masked, perception may be terminated in one of these intermediate ambiguous states. If the task is to identify the briefly presented stimulus from two options, accuracy on a single trial depends on whether perception terminates in a state that is ambiguous between the two response options: If it is ambiguous, then guessing must occur. If it is unambiguous, then the information identifies the correct option.

Figure 1 summarizes the core processes of the model for a presentation of the word "top" in the context of a two-alternative forced-choice between "top" and "cop". After an initial attentional delay, processing starts at (0,0,0); each dimension represents the processing that has occurred for a particular letter position in the presented string. When no processing has occurred, any string would be compatible with the perceptual constraints, which are summarized as "*". If the presentation ceased while perceptual processing was in this state, the participant would clearly need to guess, because the percept is consistent with the incorrect foil option as well as the correct target option, indicated in the diagram by the dashed border. Otherwise, the next perceptual event could occur for any of the three letter dimensions, moving to (1,0,0), (0,1,0) or (0,0,1); each corresponds to the relevant letter being identified without any currently useful positional information. There are two types of perceptual event for each letter dimension (0→1, which provides letter identity and minimal positional information; 1→2, which provides more detailed positional information) so perceptual processing is complete if and when (2,2,2) is reached. For the “top”-“cop” trial, of course, guessing must occur if the first letter has not been perceived, and an accurate response can be made if the first letter has been perceived (subject to the effects of masking and motor error).

Figure 2 provides a similar summary for a trial in which the stimulus presented is “pop” and the options are “pop” and “cop”. All calculations regarding the perception of each position are unchanged from the preceding example (assuming unchanged parameters). The consequences, however, differ because the first letter is now a “p” that is the same identity as the common third position of the two options, and not the “t” that occurs only in the target. As such, the perceptual constraints at (1,0,0) can be summarized as “*p*” – a “p” with no (currently) useable positional information. In other words, this “p” could be the “p” in final position of “cop” (when in fact it is a first “p” of “pop”) because to decide that this “p” is not the final “p”, it would need to be known that there is another “p”, or that there is an “o” after this “p”, or that this “p” is at the beginning – any of these additional pieces of information would disambiguate the situation. Therefore, a guess must occur for this state in the “pop”-“cop” trial, where a correct response was possible for the analogous state in the preceding “top”-“cop” example. For these examples, this is the only state whose implications differ between “top”-“cop” and “pop”-“cop” (although the situation is more complicated for internal repetitions). Since the probabilities of analogous states (those with matching co-ordinates) are the same for the two types of trial, the probability of an accurate response must be higher in the “top”-“cop” trial than the “pop”-“cop” trial.

Note that this is a qualitative prediction that does not depend on the details of the timings, probabilities, or possible movement between perceptual states. The prediction occurs because (i) for two trials of the same stimulus length, any possible perceptual state for the first trial has an analogous state for the second trial with an identical probability; (ii) there exists a state that is ambiguous between options for the second trial, but the analogous state is unambiguous for the first trial; and (iii) there is no state for which the converse is true. The states for the trials at hand meet these conditions because the first letter has been perceived and cannot be part of “cop” but can be part of “top” in the “top”-“cop” trial, but can be mistaken for the final “p” of “cop” in the “pop”-“cop” trial, because it is possible to partially perceive the stimulus in a way that there is no way to determine the position of this letter.

Similar considerations are relevant for trials in which the brief stimulus is a prime, rather than the target of an identification decision, but it is not the final state that is relevant, but rather the time spent in ambiguous states: that is, perceptual states of the prime that cannot be distinguished from the target. Lexical access occurs for words while they are consistent with the current perceptual state (what has been perceived so far), so the target is primed an amount equal to the time for which the prime percept is ambiguous with the target. (A similar process may occur for nonwords for which a representation has been established, but not nonwords in general.) Therefore, more priming occurs for prime-target combinations (in, e.g., lexical decision) that are less accurate as target-foil combinations in identification (with the brief prime of the priming trial being analogous to the brief target of the identification trial).

LTRS thus explains graded similarity-like effects with brief stimuli without positing a graded similarity-like computation as part of the cognitive processes involved. Instead, the core premise of LTRS is that such phenomena reflect the (probabilistic) incompleteness of the percept when the stimulus is brief, with potentially relevant information simply being missing. As such, what is important is how soon relevant information – such as an inconsistency between foil and target or between prime and target – can be perceived, that is, how easily an inconsistency is detected. Additional inconsistencies make it easier overall to detect that an inconsistency exists, but different information has different time courses, so the number of inconsistencies is not strictly relevant, and no single number can summarize difficulty at different time points (indeed, the ordering is not always preserved). This contrasts with other models that attribute these phenomena to a system that tolerates inconsistences and/or mistrusts the percept and have an underlying match or distance score. Although a mechanism must exist for expert readers to tolerate typographical and spelling errors, LTRS was designed to explain identification and priming phenomena without invoking this mechanism.

In the following, we detail for the original Adelman (2011) LTRS model (1) the perceptual states: their probability distributions over time, and how the perceptual constraints follow from each perceptual state; (2) calculations for predicted accuracy for the two-alternative forced-choice identification task; (3) calculations for predicted response time priming for tasks with briefly presented primes; and (4) the role of stochastic dominance in qualitative predictions of the model. Further elaborations of the model will be discussed later for Experiment 1 – (§3.3.1) where an alternative guessing rule is considered – and Experiment 4 – (§4.3.1) where the same-different identification task requires a new decision rule.

2.1.1. Perceptual processing and inference common to all paradigms

Each letter in a briefly presented string can be in one of three perceptual states: unperceived; perceived with relative location information; or perceived with full location information. On each trial, each letter starts in the unperceived state (state 0) and at times determined by a random distribution can pass to a partially perceived state (state 1) and at times determined by another random distribution from that state to a fully perceived state (state 2). These processes occur independently, so at any time during the stimulus, or at stimulus offset, any combination of different states of different letters is possible. However, each letter's state proceeds in the increasing direction (except for a possible post-mask effect). The perceptual state of the letters in an n-letter long stimulus at time t therefore can be characterized as a vector: $\boldsymbol{x}(t) = (x_1, x_2, , x_n)$ where the x_i can take values 0 (no information), 1 (partial information) or 2 (full information).

The perceptual information that can be extracted from the stimulus s in perceptual state $\boldsymbol{x}$ is complied into perceptual constraints ψ that can be summarized in a textual pattern format. We justify the value of ψ for a given $\boldsymbol{x}$ and s based on the information contained in an intermediate representation $\boldsymbol{y}$ (though the intermediate representation need never be computed to derive predictions); example values of $\boldsymbol{y}$ can be seen in Tables 1 and 2, and their interpretation will now be described.

The perceptual information that is derived so far from a particular stimulus s in a particular perceptual state is a set of perceived letters (those letters with state $x_i > 0$) each of which is tagged with some location information (whose values are only meaningful in relation to the locations of other letters), plus always available are the left and right edge co-ordinates ($l_{\min}$ and $r_{\max}$) of the entire stimulus. This information is a vector $\boldsymbol{y}$ where the i indices from $\boldsymbol{x}$ are not available, and the letters have indices j that are arbitrary and not related to letter order. A letter y_j corresponding to state 1 has the information (z_j, c_j) where z_j is the letter's identity, and c_j is a horizontal co-ordinate somewhere inside that letter; such a location is insufficient to determine if this letter is adjacent to another letter, or is first or last in the string. This is, however, sufficient information to determine the order of two perceived letters. A letter y_i corresponding to state 2 has the information (z_j, l_j, r_j) where the z_j is the letter's identity (same as state 1), l_j is its left edge co-ordinate and r_j is its right edge co-ordinate; this location information is sufficient to determine (in combination with $l_{\min}$ and $r_{\max}$ that are always known) if the letter is first or last, and if the letter is adjacent to another letter, if the edges of that letter are also available.

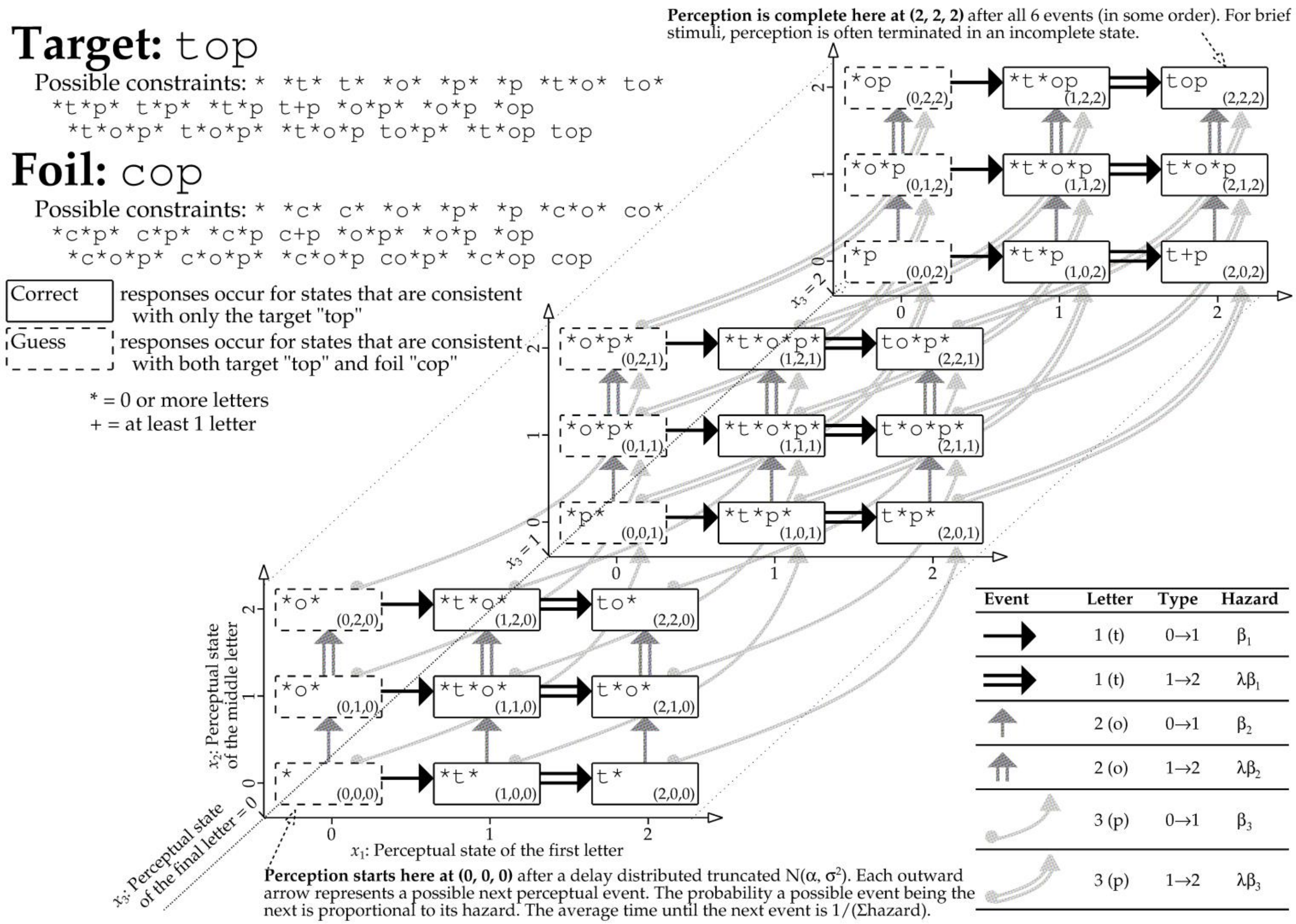

Figure 1. Diagram of all possible perceptual states (and all possible transitions between them) in LTRS when the stimulus s = “top”, and the consequences for such an identification trial with the foil “cop”. A similar diagram for a repeated-letter stimulus (“pop”) is given in Figure 2.

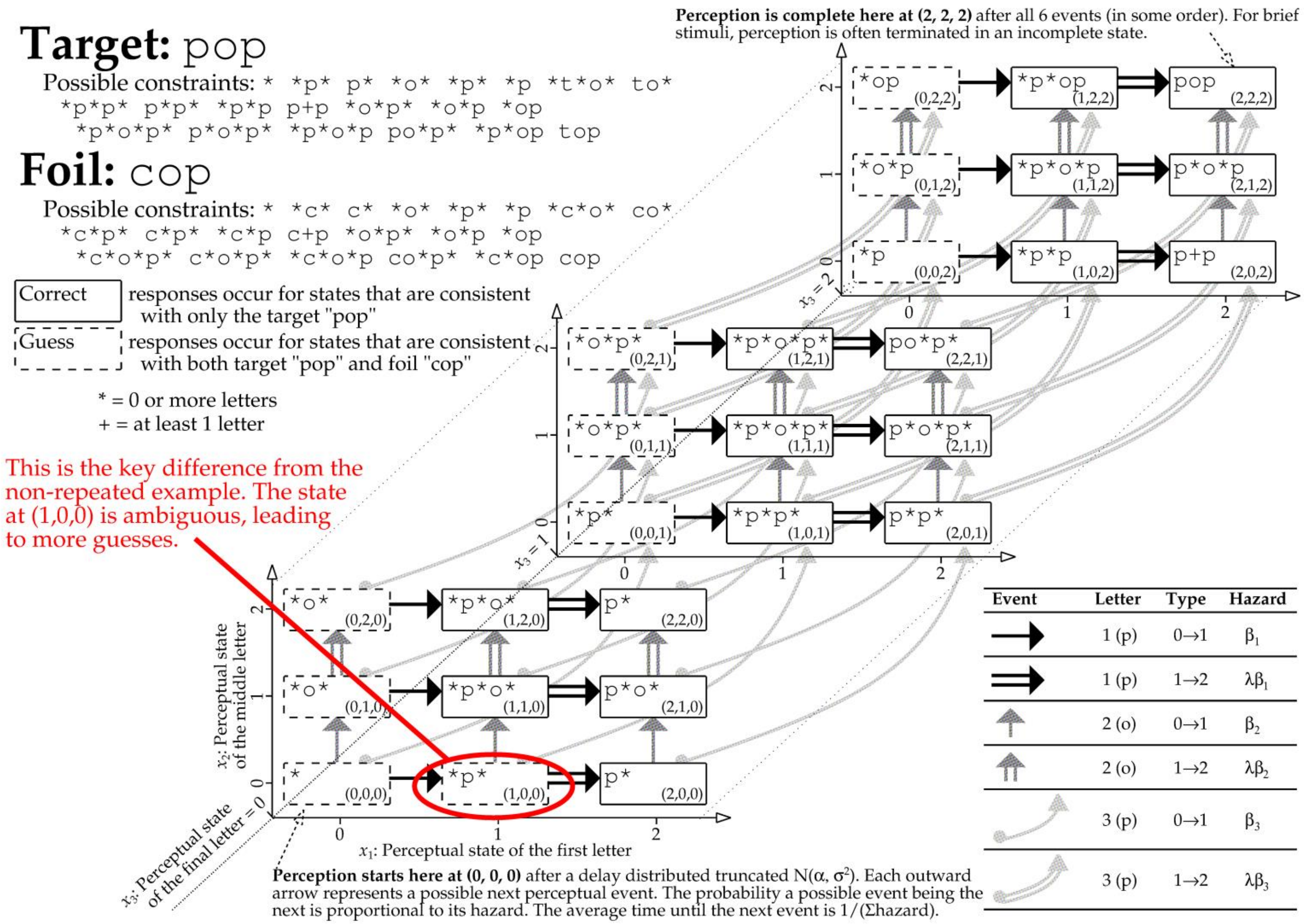

Figure 2. Diagram of all possible perceptual states (and all possible transitions between them) in LTRS when the stimulus s = "pop", and the consequences for such an identification trial with the foil "cop". A similar diagram for a non-repeated-letter stimulus ("top") is given in Figure 1.

Table 1. Enumeration of possible perceptual states $\boldsymbol{x}$, example corresponding perceptual information $\boldsymbol{y}$, and corresponding textual constraint patterns ψ for a stimulus s = "top" whose edges are at 0.4 and 4: "t" has width 1; "o" has width 1.2, and "p" has width 1.4. On any given trial, at most seven of these states are reached before post-masking. Table 2 shows a similar trial in which s = "pop".

State $\boldsymbol{x}$ = (x_1, x_2, x_3)	Information $\boldsymbol{y}$: The numbers and letter order here are examples only and would vary from trial to trial	Constraint pattern ψ	Could be "cop"?	Possible Next States
0, 0, 0	{ l_{min} = 0.4, r_{max} = 4 }	*	Y	(0,0,1) OR (0,1,0) OR (1,0,0)
0, 0, 1	{ (p, 3.1), l_{min} = 0.4, r_{max} = 4 }	*p*	Y	(0,0,2) OR (0,1,1) OR (1,0,1)
0, 0, 2	{ (p, 2.6, 4), l_{min} = 0.4, r_{max} = 4 }	*p	Y	(0,1,2) OR (1,0,2)
0, 1, 0	{ (o, 1.9), l_{min} = 0.4, r_{max} = 4 }	*o*	Y	(0,1,1) OR (0,2,0) OR (1,1,0)
0, 1, 1	{ (p, 3.1), (o, 1.9), l_{min} = 0.4, r_{max} = 4 }	*o*p*	Y	(0,1,2) OR (0,2,1) OR (1,1,1)
0, 1, 2	{ (o, 1.9), (p, 2.6, 4), l_{min} = 0.4, r_{max} = 4 }	*o*p	Y	(0,2,2) OR (1,1,2)
0, 2, 0	{ (o, 1.4, 2.6), l_{min} = 0.4, r_{max} = 4 }	+o+	Y	(0,2,1) OR (1,2,0)
0, 2, 1	{ (p, 3.1), (o, 1.4, 2.6), l_{min} = 0.4, r_{max} = 4 }	+o*p*	Y	(0,2,2) OR (1,2,1)
0, 2, 2	{ (p, 2.6, 4), (o, 1.4, 2.6), l_{min} = 0.4, r_{max} = 4 }	+op*	Y	(1,2,2)
1, 0, 0	{ (t, 0.8), l_{min} = 0.4, r_{max} = 4 }	*t*	N	(1,0,1) OR (1,1,0) OR (2,0,0)
1, 0, 1	{ (t, 0.8), (p, 3.1), l_{min} = 0.4, r_{max} = 4 }	*t*p*	N	(1,0,2) OR (1,1,1) OR (2,0,1)
1, 0, 2	{ (t, 0.8), (p, 2.6,4), l_{min} = 0.4, r_{max} = 4 }	*t*p	N	(1,1,2) OR (2,0,2)
1, 1, 0	{ (t, 0.8), (o, 1.9), l_{min} = 0.4, r_{max} = 4 }	*t*o*	N	(1,1,1) OR (1,2,0) OR (2,1,0)
1, 1, 1	{ (t, 0.8), (p, 3.1), (o, 1.9), l_{min} = 0.4, r_{max} = 4 }	*t*o*p*	N	(1,1,2) OR (1,2,1) OR (2,1,1)
1, 1, 2	{ (t, 0.8), (p, 2.6, 4), (o, 1.9), l_{min} = 0.4, r_{max} = 4 }	*t*o*p	N	(1,2,2) OR (2,1,2)
1, 2, 0	{ (t, 0.8), (o, 1.4, 2.6), l_{min} = 0.4, r_{max} = 4 }	*t*o+	N	(1,2,1) OR (2,2,0)
1, 2, 1	{ (t, 0.8), (p, 3.1), (o, 1.4, 2.6), l_{min} = 0.4, r_{max} = 4 }	*t*o*p*	N	(1,2,2) OR (2,2,1)
1, 2, 2	{ (t, 0.8), (p, 2.6, 4), (o, 1.4, 2.6), l_{min} = 0.4, r_{max} = 4 }	*t*op	N	(2,2,2)
2, 0, 0	{ (t, 0.4, 1.4), l_{min} = 0.4, r_{max} = 4 }	t*	N	(2,0,1) OR (2,1,0)
2, 0, 1	{ (t, 0.4, 1.4), (p, 3.1), l_{min} = 0.4, r_{max} = 4 }	t*p*	N	(2,0,2) OR (2,1,1)
2, 0, 2	{ (t, 0.4, 1.4), (p, 2.6, 4), l_{min} = 0.4, r_{max} = 4 }	t+p	N	(2,1,2)
2, 1, 0	{ (t, 0.4, 1.4), (o, 1.9), l_{min} = 0.4, r_{max} = 4 }	t*o*	N	(2,1,1) OR (2,2,0)
2, 1, 1	{ (t, 0.4, 1.4), (p, 3.1), (o, 1.9), l_{min} = 0.4, r_{max} = 4 }	t*o*p*	N	(2,1,2) OR (2,2,1)
2, 1, 2	{ (t, 0.4, 1.4), (p, 2.6, 4), (o, 1.9), l_{min} = 0.4, r_{max} = 4 }	t*o*p	N	(2,2,2)
2, 2, 0	{ (t, 0.4, 1.4), (o, 1.4, 2.6), l_{min} = 0.4, r_{max} = 4 }	to+	N	(2,2,1)
2, 2, 1	{ (t, 0.4, 1.4), (p, 3.1), (o, 1.4, 2.6), l_{min} = 0.4, r_{max} = 4 }	to*p*	N	(2,2,2)
2, 2, 2	{ (t, 0.4, 1.4), (p, 2.6, 4), (o, 1.4, 2.6), l_{min} = 0.4, r_{max} = 4 }	top	N	

Note – In $\boldsymbol{y}$, the ordering of letter elements is arbitrary (and ignored), and the precise numerical values vary randomly and with the properties of the physical display, but this information is abstracted away to produce location and font invariances in the constraints. The possible ψ constraints in the final column constitute C("top") the constraint list consistent with "top".

Table 2. Enumeration of possible perceptual states $\boldsymbol{x}$, example corresponding perceptual information $\boldsymbol{y}$, and corresponding textual constraint patterns ψ for a stimulus s = "pop" whose edges are at 0 and 4: "o" has width 1.2, and both "p"s have width 1.4. On any given trial, at most seven of these states are reached before post-masking. Table 1 shows a similar trial in which s = "top".

State $\boldsymbol{x}$ = (x_1, x_2, x_3)	Information $\boldsymbol{y}$: The numbers and letter order here are examples only and would vary from trial to trial	Constraint pattern ψ	Could be "cop"?	Possible Next States
0, 0, 0	{ l_{min} = 0, r_{max} = 4 }	*	Y	(0,0,1) OR (0,1,0) OR (1,0,0)
0, 0, 1	{ (p, 3.1), l_{min} = 0, r_{max} = 4 }	*p*	Y	(0,0,2) OR (0,1,1) OR (1,0,1)
0, 0, 2	{ (p, 2.6, 4), l_{min} = 0, r_{max} = 4 }	*p	Y	(0,1,2) OR (1,0,2)
0, 1, 0	{ (o, 1.9), l_{min} = 0, r_{max} = 4 }	*o*	Y	(0,1,1) OR (0,2,0) OR (1,1,0)
0, 1, 1	{ (p, 3.1), (o, 1.9), l_{min} = 0, r_{max} = 4 }	*o*p*	Y	(0,1,2) OR (0,2,1) OR (1,1,1)
0, 1, 2	{ (o, 1.9), (p, 2.6, 4), l_{min} = 0, r_{max} = 4 }	*o*p	Y	(0,2,2) OR (1,1,2)
0, 2, 0	{ (o, 1.4, 2.6), l_{min} = 0, r_{max} = 4 }	+o+	Y	(0,2,1) OR (1,2,0)
0, 2, 1	{ (p, 3.1), (o, 1.4, 2.6), l_{min} = 0, r_{max} = 4 }	+o*p*	Y	(0,2,2) OR (1,2,1)
0, 2, 2	{ (p, 2.6, 4), (o, 1.4, 2.6), l_{min} = 0, r_{max} = 4 }	+op*	Y	(1,2,2)
1, 0, 0	{ (p, 0.8), l_{min} = 0, r_{max} = 4 }	*p*	Y	(1,0,1) OR (1,1,0) OR (2,0,0)
1, 0, 1	{ (p, 0.8), (p, 3.1), l_{min} = 0, r_{max} = 4 }	*p*p*	N	(1,0,2) OR (1,1,1) OR (2,0,1)
1, 0, 2	{ (p, 0.8), (p, 2.6,4), l_{min} = 0, r_{max} = 4 }	*p*p	N	(1,1,2) OR (2,0,2)
1, 1, 0	{ (p, 0.8), (o, 1.9), l_{min} = 0, r_{max} = 4 }	*p*o*	N	(1,1,1) OR (1,2,0) OR (2,1,0)
1, 1, 1	{ (p, 0.8), (p, 3.1), (o, 1.9), l_{min} = 0, r_{max} = 4 }	*p*o*p*	N	(1,1,2) OR (1,2,1) OR (2,1,1)
1, 1, 2	{ (p, 0.8), (p, 2.6, 4), (o, 1.9), l_{min} = 0, r_{max} = 4 }	*p*o*p	N	(1,2,2) OR (2,1,2)
1, 2, 0	{ (p, 0.8), (o, 1.4, 2.6), l_{min} = 0, r_{max} = 4 }	*p*o+	N	(1,2,1) OR (2,2,0)
1, 2, 1	{ (p, 0.8), (p, 3.1), (o, 1.4, 2.6), l_{min} = 0, r_{max} = 4 }	*p*o*p*	N	(1,2,2) OR (2,2,1)
1, 2, 2	{ (p, 0.8), (p, 2.6, 4), (o, 1.4, 2.6), l_{min} = 0, r_{max} = 4 }	*p*op	N	(2,2,2)
2, 0, 0	{ (p, 0, 1.4), l_{min} = 0, r_{max} = 4 }	p*	N	(2,0,1) OR (2,1,0)
2, 0, 1	{ (p, 0, 1.4), (p, 3.1), l_{min} = 0, r_{max} = 4 }	p*p*	N	(2,0,2) OR (2,1,1)
2, 0, 2	{ (p, 0, 1.4), (p, 2.6, 4), l_{min} = 0, r_{max} = 4 }	p+p	N	(2,1,2)
2, 1, 0	{ (p, 0, 1.4), (o, 1.9), l_{min} = 0, r_{max} = 4 }	p*o*	N	(2,1,1) OR (2,2,0)
2, 1, 1	{ (p, 0, 1.4), (p, 3.1), (o, 1.9), l_{min} = 0, r_{max} = 4 }	p*o*p*	N	(2,1,2) OR (2,2,1)
2, 1, 2	{ (p, 0, 1.4), (p, 2.6, 4), (o, 1.9), l_{min} = 0, r_{max} = 4 }	p*o*p	N	(2,2,2)
2, 2, 0	{ (p, 0, 1.4), (o, 1.4, 2.6), l_{min} = 0, r_{max} = 4 }	po+	N	(2,2,1)
2, 2, 1	{ (p, 0, 1.4), (p, 3.1), (o, 1.4, 2.6), l_{min} = 0, r_{max} = 4 }	po*p*	N	(2,2,2)
2, 2, 2	{ (p, 0, 1.4), (p, 2.6, 4), (o, 1.4, 2.6), l_{min} = 0, r_{max} = 4 }	pop	N	

Note – In $\boldsymbol{y}$, the ordering of letter elements is arbitrary (and ignored), and the precise numerical values vary randomly and with the properties of the physical display, but this information is abstracted away to produce location and font invariances in the constraints. The possible ψ constraints in the final column constitute C("pop") the constraint list consistent with "pop".

The useful perceptual information present is more conveniently summarized as the textual pattern ψ, which can include wildcards. Two wildcard characters are useful to express constraints in LTRS's perceptual information: We use + to indicate that at least one letter is needed between the preceding and following item for a match to be made and * to indicate that letters may be present or absent between the preceding and following item. In this format, a whole pattern match is implied, and the beginning and end must match. Thus, go matches g*o and *g*o but not g+o; gro and grio match all three; and agro matches *g*o but not g*o or g+o; and grob matches none of them. Table 1 illustrates the relationship between $\boldsymbol{x}$, $\boldsymbol{y}$ and ψ for the example stimulus "top", adding $\boldsymbol{y}$ examples to the information of Figure 1; and Table 2 illustrates the same for the example "pop" of Figure 2.

For a perceptual state $\boldsymbol{x}(t)$ at a time t during perception of a stimulus s, the perceptual system has available perceptual information $\boldsymbol{y}(\boldsymbol{x}(t); s)$ from which it can compute the perceptual constraints ψ by a procedure equivalent to that in Table 3.

Table 3. Pseudocode algorithm to summarize $\boldsymbol{y}$ constraints as a textual pattern, ψ.

```
Identify the left-most unprocessed letter y_j with the lowest left-edge l_j or internal position c_j
If l_j is unavailable or l_j ≠ l_min (i.e, is not at the stimulus left-edge), append "*" to pattern
Append the current letter identity z_j to pattern.
Mark the current letter y_j as processed.
Keep a temporary record of the most recently processed letter p = j.
While elements of y remain unprocessed:
        Update j by identifying the left-most unprocessed letter y_j – that with the lowest l_j or c_j
        If the previous letter's right-edge r_p or the current letter's left edge l_j is unavailable, then append "*" to
        pattern
        If both r_p and l_j are available and the current letter is not adjacent to the previous letter, i.e., r_p ≠ l_j, then
        append "+" to pattern
        Append the current letter identity z_j to pattern
        Mark current letter y_j as processed
        Record the most recently processed letter p = j.
Loop
If the right edge of the last letter r_j is unavailable or r_j ≠ r_max (i.e., it is not at the right edge of the stimulus), then
append "*" to pattern
```

The constraints ψ ultimately follow from the following: If a letter is in state 1 or 2, its identity is known. If a letter is in state 2, it is also known whether it is the initial or final letter. If two letters are both in either state 1 or 2, their order is additionally known. If two letters are both in state 2, then whether they are adjacent is also known (non-adjacency can also be implied by an intervening letter).

When we are calculating the predictions of the model, it is unnecessary to do computations that involve $\boldsymbol{y}$. We can work out ψ directly from $\boldsymbol{x}$ – through the procedure in Table 4 – and this simplifies away from the details of positional information (c_j, l_j, r_j, l_{min}, r_{max}), as the result is the same as would be achieved via obtaining $\boldsymbol{y}$ and applying Table 3's procedure.

Table 4. Pseudocode algorithm to summarize perceptual constraints arising from a given $\boldsymbol{x}$ and s as a textual pattern, ψ.

```
If x_1 < 2 (the first position does not have full positional information), append "*" to pattern
If x_1 > 0 (the first position has been perceived), append s_1 (its identity) to pattern
If x_1 = 2, let a = 1
Let i = 2; Do:
        If x_i = 2 and a = 1 and x_{i-1} ≠ 2, append "+" to pattern
        If x_i = 0 and x_{i-1} > 0, append "*" to pattern
        If x_i > 0 append s_i to pattern and let a = 0
        If x_i = 2, let a = 1
        Increment i
While i ≤ n
If x_n = 1, append "*" to pattern
```

For any given vocabulary item or response option w, only a finite set of sets of perceptual constraint patterns can be reached from the 3^n possible $\boldsymbol{x}$ states; this set of consistent patterns $C(w)$ can be computed by iterating over the 3^n possible $\boldsymbol{x}$ states. The perceptual information from a stimulus s at a given time t, $\mathbf{y}(t; s)$ on any trial can be compared to the set $C(w)$ to determine whether the percept is consistent with the vocabulary item or response option w. The Boolean values, $v(w; \boldsymbol{y}) = 1$ if $\psi(\boldsymbol{y}) \in C(w)$, 0 otherwise, are central to the LTRS conception of how perceptual information is used: Incomplete perceived perceptual information is compared to the incomplete perceptual information that could possibly be perceived from a stimulus, and modus tollens applies to determine inconsistency. That is: If the stimulus s were the option w, then the percept ψ would be one of the $C(w)$; if it is not, then s cannot be w. So, for the example stimulus "top", C("top") is exactly the elements of the third column of Table 1. When deciding if a percept might be "top", it is compared to the possible percepts of "top", C("top").

To produce numerical predictions, specification of the probabilistic time course of the percept $\boldsymbol{x}$ (in Equations 2-4) makes it possible to compute (by summing the probability of all consistent cases) the probability that the percept from a stimulus s at time t will be consistent with a vocabulary item or response option w, that is:

$$P(V(w) = 1; s, t) = \sum_{\boldsymbol{x}:\, \psi(\boldsymbol{x};\, s) \in C(w)} P(\boldsymbol{X}(t) = \boldsymbol{x}). \tag{1}$$

a value used to compute all LTRS predictions. That is, the probability that the percept $\boldsymbol{x}$ is consistent with the item w under consideration (that $v(w)$ is 1) is the probability that the percept is in one of the states that produces a constraint pattern ψ that is among the constraint patterns consistent with the considered item w ($C(w)$). Since these states are mutually exclusive, the total probability of consistency is the sum of the probabilities of these consistent states (those states $\boldsymbol{x}$ whose corresponding pattern ψ is in $C(w)$).

Consider our examples in Table 1 and 2 of targets "top" and "pop" with respect to the foil "cop". The difference in probability of ambiguity comes from the different terms involved in the sum: $\boldsymbol{x}: \psi(\boldsymbol{x};\ s) \in\ C(w)$ for the two different targets. For the trial with the target s = "pop" and foil w = "cop", the sum is based on the first ten $\boldsymbol{x}$ values in Table 2. In contrast, for the target s = "top" with the same foil, only the first nine of these are included from Table 1. So long as the probability of the $\boldsymbol{x}$ = (1,0,0) state (the tenth row) is greater than 0, $P(V(\text{cop}) = 1)$ – the probability the foil "cop" is compatible with the percept – is higher for "pop"-"cop" than it is for "top"-"cop". Percepts arising from "pop" are more probably ambiguous than those arising from "top," and this does not depend on the specifics of the timing or sequencing of the perceptual states, nor any parameters (so long as the same parameters are used in both cases).

The above aspects of the model serve to make qualitative predictions in many (but not all) cases of interest (so long as it is assumed certain probabilities will be greater than 0, which is true for finite parameter values). To make specific quantitative predictions, the probabilities of perceptual states over the time course of perception need to be specified. The time course of the percept $\boldsymbol{x}$ has two components, an initial random perceptual delay t_0 (in milliseconds) affecting all components equally on the same trial, and the otherwise independent subsequent transitions from 0 to 1 to 2 in the individual elements $\boldsymbol{x}_i$. The transitions from 0 to 1 have a constant hazard $\beta_{i|n}$ dependent on the letter position i and stimulus length n; when time is expressed in milliseconds, the β values have a natural unit of kHz (letters/ms). The transitions from 1 to 2 have a constant hazard $\lambda\beta_{i|n}$ proportional to that for the first transition (λ is a ratio of kHz/kHz). The specification of constant hazard defines an exponential distribution, so it can be shown that:

$$P(X_i = 0 \mid T_0 = t_0;\ t) = e^{-\beta_{i|n}(t-t_0)};$$

$$P(X_i = 2 \mid T_0 = t_0;\ t) = 1 - \frac{1}{\lambda - 1}\left(e^{-\beta_{i|n}(t-t_0)} - e^{-\lambda\beta_{i|n}(t-t_0)}\right) - e^{-\beta_{i|n}(t-t_0)} \quad \text{if } \lambda \neq 1;$$

$$P(X_i = 2 \mid T_0 = t_0;\ t) = 1 - \left(\beta_{i|n}(t - t_0) + 1\right)e^{-\beta_{i|n}(t-t_0)} \quad \text{if } \lambda = 1; \text{ and}$$

$$P(X_i = 1 \mid T_0 = t_0;\ t) = 1 - P(X_i = 0 \mid T_0 = t_0;\ t) - P(X_i = 2 \mid T_0 = t_0;\ t). \quad (2)$$

These processes are independent, so

$$P(\boldsymbol{X} = \boldsymbol{x} \mid T_0 = t_0;\ t) = \prod_i P(X_i = x_i \mid T_0 = t_0;\ t)\,. \quad (3)$$

The distribution of T_0 is $N(\alpha, \sigma)$ (with α and σ normally in milliseconds) so the non-conditional probabilities for $V = 0$ and (complementarily) $V = 1$ are computed by numerical integration of

$$P(V(w) = 0; s, t) = \int \sum_{\boldsymbol{x}:\, \psi(\boldsymbol{x};s) \notin C(w)} \frac{dP(T_0 = t_0)}{dt_0} \prod_i P(X_i = x_i \mid T_0 = t_0;\ t)\ dt_0 \quad (4)$$

where the dP/dt_0 is normal density function.

2.1.2. Two-alternative forced choice perceptual identification

In the task of Experiment 1, a stimulus *s* is presented for a time *t* then masked, and correct option *s* and a foil option *f* are presented as a two-alternative forced choice, so if only one *v* is 1, then that option should be chosen, if possible. Clearly, $v(s)$ is always 1 because the percept is always consistent with the presented stimulus. Therefore, accuracy only depends on the probability of $v(f)$ being 0 after the post-mask – that is, how probably the percept rules out the foil by providing constraints inconsistent with the foil.

The effect of the post-mask can be to cause a loss of information for a letter. Such a transition back to 0 can occur for each x_i independently with probability φ. The $P(X_i = 1)$ and $P(X_i = 2)$ of Equation 2 are then adjusted to masked versions of these probabilities by multiplication by (1 -

φ) and the lost probability is added to $P(X_i = 0)$. These probabilities for the masked case can be inserted into Equations 3 and 4 to calculate $P(V = 0)$ specifically for the masked case.

So, if after the effect of the post-mask, $v(f) = 0$, then responding is correct, unless a premature guess occurs with probability ε. If $v(f) = 1$, then a guess must be made. Accuracy is therefore computed thus:

$P(\text{no guess}) = (1 - \varepsilon)P(V(f) = 0; s, t, \text{masked})$

$P(\text{accurate}) = P(\text{no guess}) + g_c(s, f)P(\text{guess}) =$

$$(1 - \varepsilon)P(V(f) = 0; s, t, \text{masked}) + g_c(s, f)\big(1 - (1 - \varepsilon)P(V(f) = 0; s, t, \text{masked})\big) \quad (5)$$

where $g_c(s, f)$ is the probability of a correct guess.

In most previous applications of LTRS to this task, trials in which w_1 was the stimulus and w_2 was the foil formed the same condition as trials in which w_2 was the stimulus and w_1 was the foil, so that the probability of responding correctly from a guess averaged to $g_c = 0.5$ within a condition. Some forms of bias could occur but would not affect the overall accuracy of the conditions: A bias towards, for instance, the more frequent of an option pair would average out between the trials in which that response was correct and those in which it was incorrect. Moreover, there is no way one could be biased towards, for instance, the "transposed-letter" option in a pair such as swan-sawn, because this is a reciprocal relationship within the pair. (A hand bias was fitted in some cases, but the fitting for the current experiments did not require this level of detail.)

This version of the model was the one we first fitted to Experiment 1. We also describe below in the information on modelling individual experiments (i) a modified version with a different guessing rule for Experiment 1; and (ii) a similar model for the same-different task of Experiment 4.

2.1.3. Masked priming

In priming paradigms, a briefly presented prime *s* is presented before a task must be performed on a target *w* that has an established representation. In prior applications of LTRS, *w* was a word with an established lexical entry in the lexical decision, and no priming was predicted for "unseen" nonwords. In the present Experiments 2 and 3, a representation w_0 is established for a reference nonword stimulus presented clearly at the beginning of the trial. The task is a same-different judgment between w_0 and w, in which s typically has a priming effect on trials where the correct answer is same. (This difference between seen [same] and unseen [different] targets is analogous to the word-nonword distinction in lexical decision.) This generalization is simple in LTRS because it does not specify how other aspects of lexical processing and the actual lexical decision occur, only how much time is later "saved" due to prime processing.

In LTRS, priming is caused by the prime going through incompletely perceived states that are consistent with the target, so a non-identical prime percept can be ambiguous with the target; that is $v(w)$ can be 1 during the prime even if the prime is not identical to the target. Priming is not produced by only the ambiguity at the offset of the prime. Rather, potential identities for the prime are undergoing processing when their v is 1, and the time spent on this processing of the target during the prime is saved on later processing of the target, affecting response times. Prime processing is assumed to continue after prime offset for a period lasting ω (omega) during which no changes occur to the prime percept.

Priming as access for a target w from prime s on a trial with prime duration l is therefore

$$\gamma = \int_{t=t_0}^{l} v(w; \boldsymbol{y}(t; s))\, dt + \omega v(w; \boldsymbol{y}(l; s))$$

Mean priming as access for a fixed t_0 is therefore

$$E(\Gamma \mid s, w, l, t_0) = \int_{t=t_0}^{l} P(V(w) = 1; s, t)\, dt + \omega P(V(w) = 1; s, l)$$

The dependency on t_0 is removed by observing that priming must always be empirically measured relative to another condition, and $\int_{t=0}^{t_0} v(w)\, dt$ does not depend on s. The predicted mean relative priming of a target w by a prime s_1 relative to its control s_0 is therefore just the difference between the mean priming for each of the two primes (after adding $\int_{t=0}^{t_0} P(V(w) = 1; s, t)\, dt$ to each):

$$\begin{aligned} &E(\Gamma \mid s_1, w, l) - E(\Gamma \mid s_0, w, l) \\ &\qquad = \int_{t=0}^{l} P(V(w) = 1;\ s_1, t)\, dt + \omega P(V(w) = 1;\ s_1, l) \\ &\qquad - \left(\int_{t=0}^{l} P(V(w) = 1;\ s_0, t)\, dt + \omega P(V(w) = 1; s_0, l)\right) \qquad (6) \end{aligned}$$

whose integrals do not depend on t_0 and can be evaluated by numerical integration.

Explicit modelling of any final judgment regarding stimuli that are not presented briefly is beyond the scope of LTRS; only the priming effect of the briefly presented stimulus is modelled.

2.1.4. Identifying stochastic dominance and stochastic equivalence to establish parameter-free LTRS predictions

It is possible in several cases to establish qualitative parameter-free predictions of LTRS by pairing (or grouping) analogous states that have the same probability across two (or more) conditions.

For the same parameters and number of letters in the briefly presented stimulus, the probability of each possible state $\boldsymbol{x}$ does not depend on the properties of the stimuli in the experiment. For each condition, we can examine the v for the relevant stimulus-probe pairing (target-foil [remembering v is always 1 between target and target] or prime-target) for each $\boldsymbol{x}$.

If for all $\boldsymbol{x}$, the two vs (one for each condition) are equal, then stochastic equivalence implies that the accuracy or priming must be the same between the two conditions, the predicted accuracy will be the same in two conditions, regardless of parameters. For example, in an identification task, if the stimulus is **cat** and the foil conditions are **coo** and **cob**: if the **a** or **t** is perceived, both **coo** and **cob** are inconsistent with the stimulus; otherwise, both are consistent.

If instead there are differences at some $\boldsymbol{x}$ states, and these differences are all in the same direction (among states that are different, one condition always has a v of 1), then the more-often ambiguous condition will be less accurate or produce more priming, regardless of parameters. For example, if the stimulus is **cat** and the foil conditions are **act** and **hot**, then if either the **c** or **a** is perceived, then **hot** is inconsistent, but **act** may or may not be consistent, depending on the other information perceived. The cases when **act** is inconsistent involve both **c** and **a** being perceived, the precise position of **c** (initial) being perceived, or the precise position of both **a** and **t** being perceived (**a** before **t** does not suffice, but **at** is informative). In none of these cases is **hot** consistent. Thus, regardless of parameters, **hot** is more accurately rejected than **act**.

Parameter-free qualitative predictions do not always exist, but they do for Experiments 1-3 that follow, as well as for two of the simple effects (and hence the one corresponding main effect) and the interaction of Experiment 4. A detailed enumeration of these cases is given in Appendixes C and D.

2.2. The Overlap Model

The Overlap Model (Gomez et al., 2008; Ratcliff, 1981) is a model whose defining component is a process to calculate an *overlap* score between briefly perceived strings and more stably represented strings that can be entered into a decisional model. Overlap is calculated on the basis the positional representation of letters in strings can be treated as containing uncertainly as to the location of a letter that is indexed by a probability distribution. For strings that are no longer visible (i.e., briefly presented then masked) the uncertainty of an individual letter in position *i* is indexed by the normal distribution $N(i, s_i^2)$ where s_i is a position-dependent standard deviation parameter. By contrast, for a clearly presented test alternative, there is less uncertainty, and whatever uncertainty exists is contained within a region of width 1 centered on the position *i*, i.e., $(i - ½, i + ½)$.

The calculation of overlap between two strings, *T* and *S*, is described by these authors as being:

$$\sum_{i=1}^{5} \int_{i-½}^{i+½} f_T(x)dx \int_{i-½}^{i+½} f_S(x)dx \qquad (7)$$

for 5-letter strings, alongside a diagram like Figure 3.

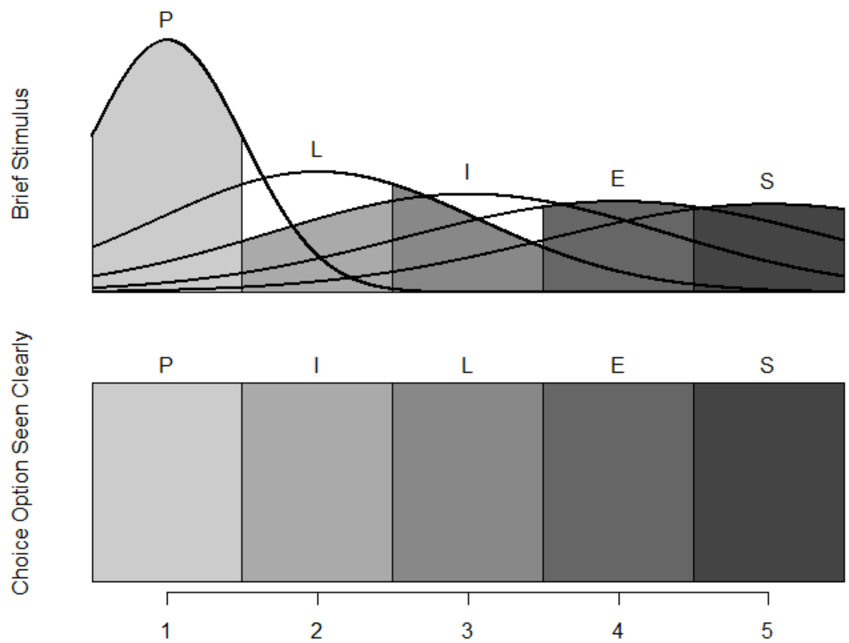

Figure 3. Typical graphical representation of the calculation of overlap for a response option “piles” when the briefly presented stimulus is “plies”.

In all applications, f_T is a density representing the visible string (target or foil) at test in the lower part of the Figure, and f_S is a density representing the uncertainly positioned brief stimulus in the upper part of the Figure. The diagram shows an f_T for each position, and an f_S in each position, and the overlap is the sum of the shaded regions of the upper portion of the diagram. It is not explicit in Equation 7 – but it is clear in the diagram and the authors' code – that it is meant that in each term of the sum it is f_T for the letter in slot i and f_S for the letter (if such exists) of the same identity. Indeed, the code involves two nested loops whose computation is more explicitly expressed using a double sum:

$$O(T,S) = \sum_{i=1}^{5}\sum_{j=1}^{5} I(i,j) \int_{i-½}^{i+½} f_{T,i}(x)dx \int_{i-½}^{i+½} f_{S,j}(x)dx \quad (8)$$

where $I(i, j)$ is 1 if the identity of letter i in T matches the identity of letter j in S, and 0 otherwise; $f_{T,i}$ is explicitly the density of letter i in string T; and $f_{S,j}$ is explicitly the density of letter j in string S.

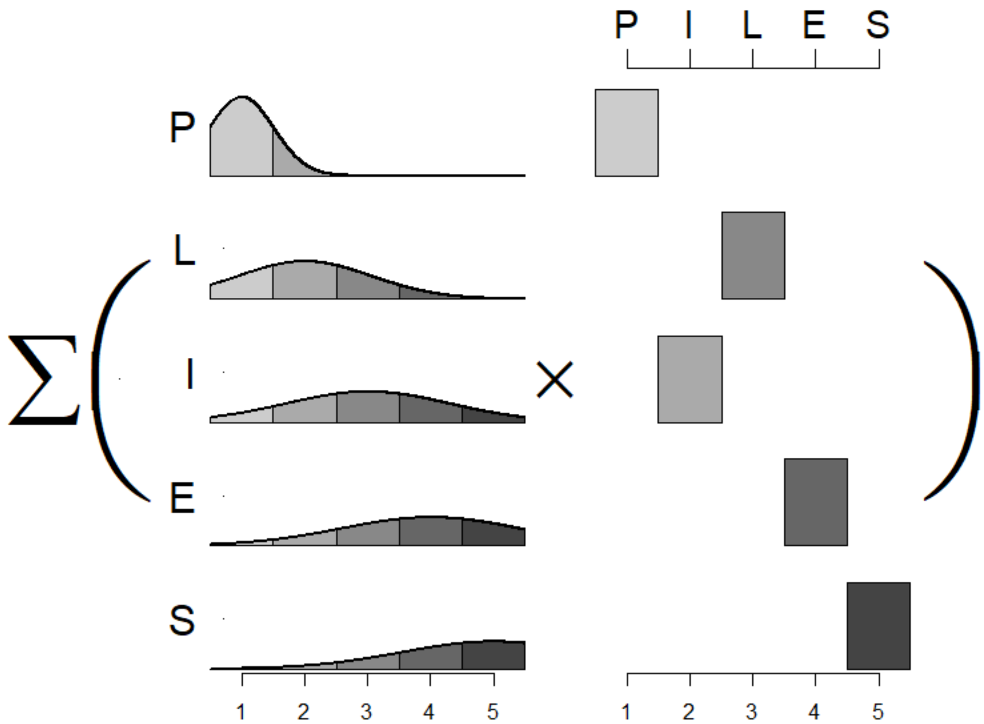

Figure 4. Expanded graphical representation of the calculation of the overlap for a response option “piles” when the briefly presented stimulus is “plies” (i.e., the response option is a transposed-letter foil). Each row represents a position (and identity) in the briefly presented stimulus, each column represents a position in the response option it is being compared with. The left curved areas represent the positional uncertainty from the briefly presented stimulus. The right rectangles represent the lack of uncertainty from the clearly visible response option, and are present (of value 1) if the letter identities match between the stimulus and response option. Elements occupying the same row and column of the left and right arrays are multiplied before summation. Since the right elements are always 1 or 0, then in effect, the right array shows which elements of the left are added together to calculate the overlap.

The calculation can be understood more clearly if we consider each of the five normal distributions for each of the briefly presented letters ($f_{S,j}$) separately, on the left hand side in each of the rows of Figure 4. Each is split into five possible regions of overlap. On the right hand side of the diagram, the identity matching is combined with the distributions of the clearly displayed option ($I(i, j).f_{T,i}$). The areas on the left are multipled by the corresponding areas on the right, and their sum is the overlap. Each area on the right is either one or zero, indicating whether or not the identities match; in the model code, this is exactly how the overlap is in fact calculated, i.e:

$$O(T,S) = \sum_{i=1}^{5}\sum_{j=1}^{5}\int_{i\,-\,½}^{i\,+\,½} f_{S,j}(x)dx\, I(i,j).$$

This representation also clearly shows how overlap calculations are made when a letter is repeated in either the briefly presented stimulus (Figure 5) or the clearly seen response option (Figure 6). In both cases, the repeated letter provides additional overlap because there is both overlap within the matching position and overlap from "leakage" of the normally distributed representation centred on one position into another position.

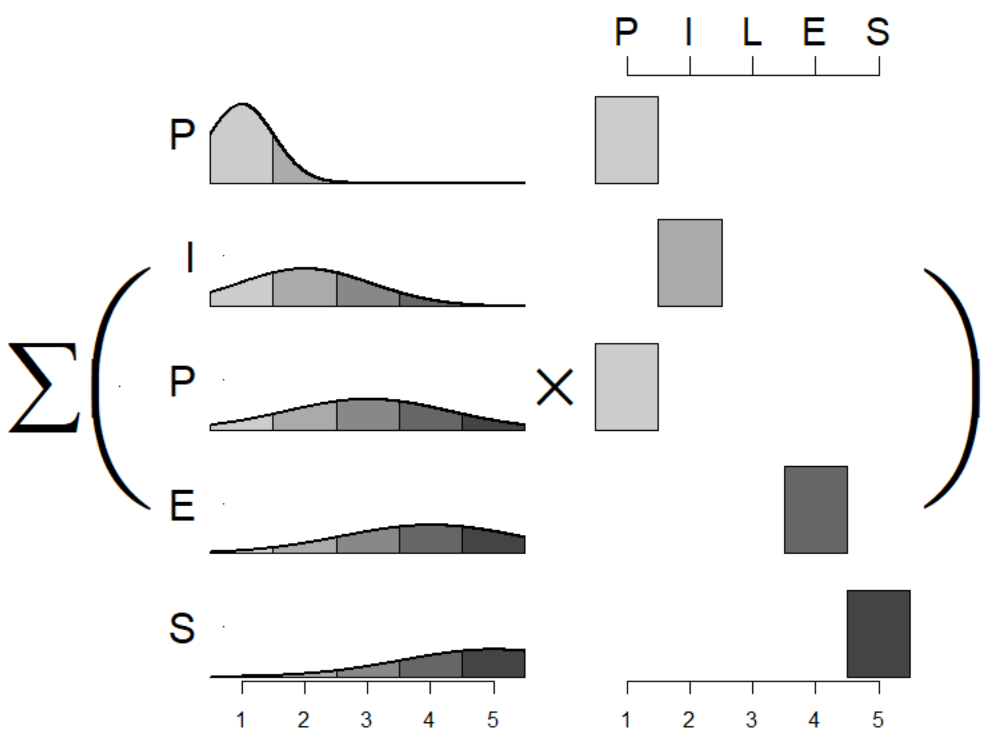

Figure 5. Expanded graphical representation of the calculation of the overlap for a response option "piles" when the briefly presented stimulus is "pipes". The interpretation of the diagram is as for Figure 4. Both instances of the repeated "p" in the stimulus contribute to the overlap, even though there is only one instance in the response option.

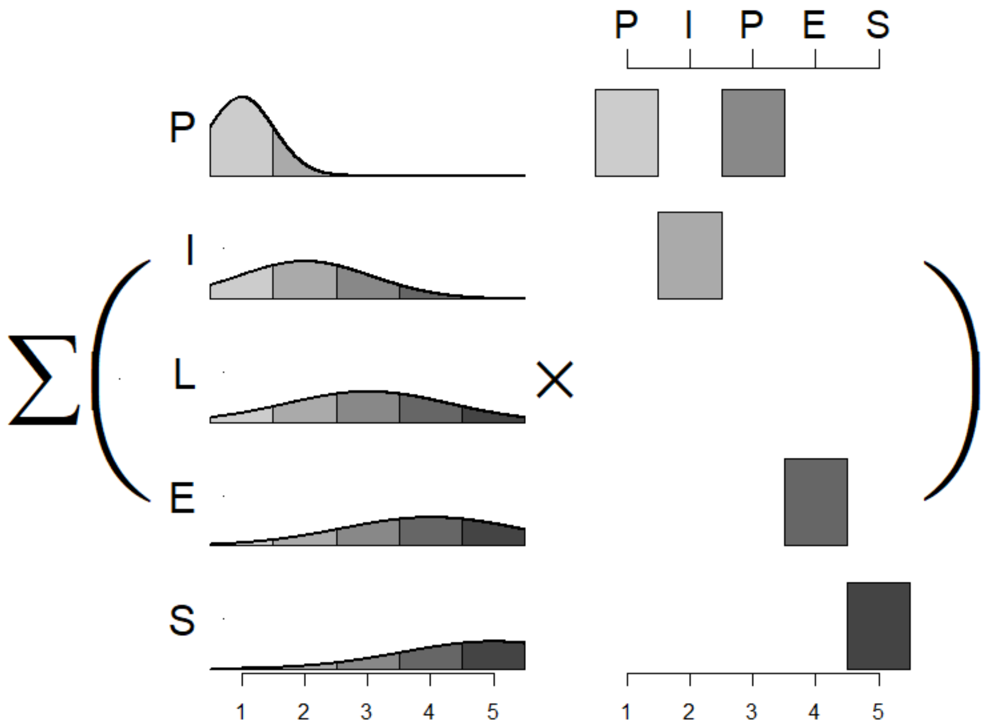

Figure 6. Expanded graphical representation of the calculation of the overlap for a response option "pipes" when the briefly presented stimulus is "piles". The interpretation of the diagram is as for Figure 4. The single instance of "p" in the stimulus makes two contributions to the overlap, as there is leakage in its positional representation to overlap with both instances of "p" in the response option.

For response alternatives that involve a deletion or insertion of a letter, resulting in an option with a different length to the stimulus, a further assumption is required. For strings of similar lengths, these comparisons involving a stretching process that is described via diagram by Gomez et al. (2008) and made explicit in code provided by Gomez (2020): The stimulus string and the test string are put on a common scale by adjusting the test string *T* to probe the region from 0.5 to *m* + 0.5, where *m* is the number of letters in *S*, while maintaining the total area of each $f_{T,j}$.

$$O(T,S) = \sum_{i=1}^{n}\sum_{j=1}^{m} I(i,j) \int_{\frac{(\mathrm{i}-1)\mathrm{m}}{\mathrm{n}}+½}^{\frac{\mathrm{im}}{\mathrm{n}}+½} f_{S,j}(x)dx$$

where there are n letters in T and m letters in S. f_T has been stretched and $f_{S,j}$ has not, so the normal distributions continue to have their original means j and variances s_j^2. This is illustrated in Figure 7, where a six-letter alternative is squeezed into the space of five letters (between 0.5 and 5.5), and the five-letter stimulus distributions are similarly partitioned into six regions instead of five.

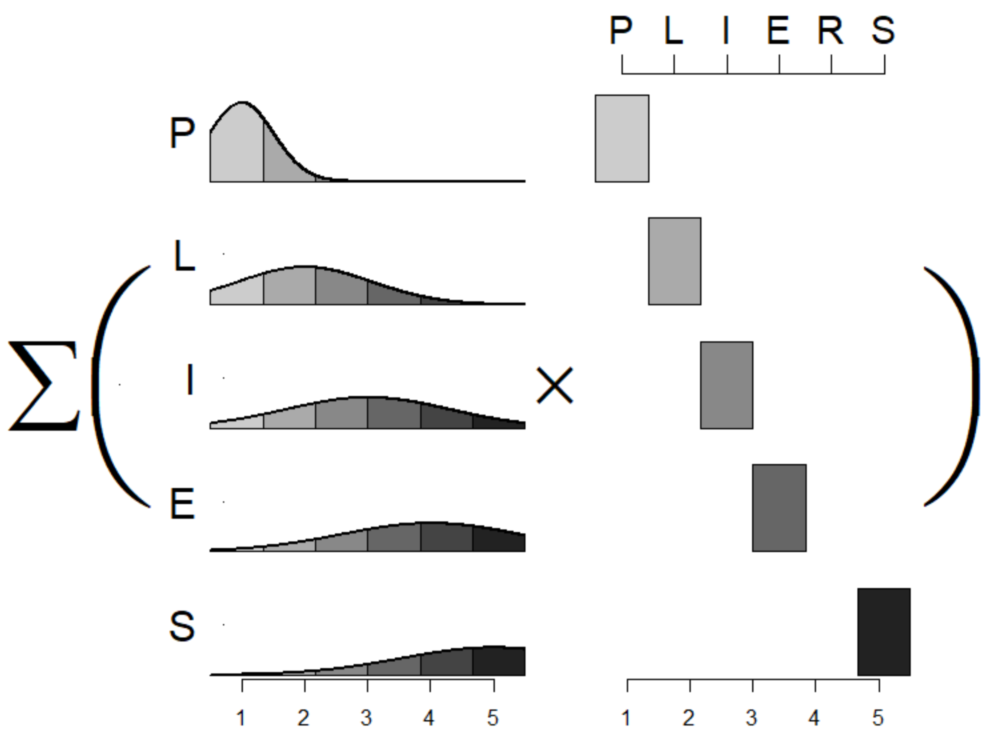

Figure 7. Expanded graphical representation of the calculation of the overlap for a response option "pliers" when the briefly presented stimulus is "plies". The interpretation of the diagram is as for Figure 4. When the lengths of stimulus and response option do not match a stretching or shrinking process occurs. In this example, while the number of columns is increased to 6 to match the number of letters, the numerical labels of the axis are still linked to the number of letters in the stimulus.

2.2.1. Constraint of standard deviation parameters

Although Gomez et al. (2008) treated the s_j parameters controlling the width of the standard deviation as free parameters in their original fitting of their experiments, they went on to also examine the fits with a constraining formula:

$$s_j = d.\left(1 - e^{-(i-0.5)/r}\right)$$

where *d* is an asymptote parameter for the largest standard deviation (for a letter infinitely to the right) and *r* is a rate parameter controlling how quickly the asymptote is reached (and how much lower the lowest standard deviation – for the initial letter – is than the asymptote). Gomez (2020) indicates that this version of the model – with fewer parmeters – should be considered the standard version.

2.2.2. Two-alternative forced choice perceptual identification

For a two-alternative forced choice task, between two possible test strings T_1 and T_2, given a briefly presented *S*, Gomez et al. (2008) modeled the decision as being made using a Choice Rule based on the overlap values raised to a power:

$$P(\text{choose } T_k) \propto O(T_k, S)^{a_k} \tag{9}$$

Although a single $a = a_k$ was fitted regardless of the identity of the stimuli in designs without repeated letters by Gomez et al. (2008), different values of a_k were used for *T*s with and without letter repetitions. A lower value was fitted for options that had repeated letters, which Gomez et al. interpreted as reflecting a bias against these options.

2.2.3. Masked priming

For masked primed lexical decision, Gomez et al. (2008) suggest it may be reasonable to examine the overlap scores $O(T, S)$ with *S* as the prime (presented briefly and not available at response) and *T* as the target of the lexical decision, and they illustrate this as a means to obtain ordinal predictions for experiments comparing different prime types with the same target stimuli (using an alternative means of dealing with comparing strings of different lengths). It is not clear how this should be generalized to comparisons between targets of different type (e.g., different lengths, containing repetitions or not) because $O(T, T)$ is not the same across different types of stimuli (unlike other forms of match score where self-similarity is 1).

3. Experiment 1: Perceptual Identification with Insertions and Deletions

In this experiment, the two-forced-choice perceptual identification task was used for the investigation of repeated letter effects. A previous study using the same task for exploring such effects was the study of Gomez et al. (2008). What sets the two studies apart is the critical manipulation used in the two studies, evident in the relationship between the target and the foil (the wrong choice). In the present study, the target and the foil were never the same length. The foil was formed by either deleting or inserting a letter. In their Experiment 4, Gomez et al. compared strings with the same length that included replacements and transpositions. They had three main conditions in which the repeated letters were either in the target, or in the foil or in both. The repeated letters in their conditions also appeared within different distances. The present study, on the other hand, focuses on a non-adjacent repetition in which the repeated letters are always separated by one letter. This choice was motivated by the stable inhibitory pattern for repeated letters within that distance demonstrated in Trifonova and Adelman (2019), as well as by the necessity for more evidence due to the gap in the orthographic processing research literature. The purpose of the experiment was to test whether the number of nonadjacent repeated letters could be determined in early perceptual stages. If the perceptual system is not able to detect or keep track of the number of the identities, such processing limitations could explain the inhibitory effect reported in the study of Trifonova and Adelman.

Another important aim of this experiment was to provide evidence regarding the cause of the low accuracy results for the repeated letter targets of Experiment 4 in Gomez et al. (2008). One possible explanation for their result could be a bias toward choosing targets with no repetition due to some unnaturalness of targets with repeated letters. Such an explanation does not imply any effects due to the presence of repeated letters per se. Another explanation, however, could be that the presence of repeated letters raises the level of processing difficulty of those targets. To test the reason of the results reported by Gomez et al. the comparison was made between trial type (repeat vs unique), rather than target type, as in the case of Gomez et al. As the repeated letters were not only in the target, but also in the foil, it could be tested whether any effect of repetition could be attributed to bias of choosing a string with no letter repetition. Choosing a deletion foil (the string with no repeated letters) will lead to lower accuracy in repeated trial condition with eight-letter targets. However, choosing seven-letter target (the string with no repeated letters) more often than the insertion foil (with repeated letter) will lead to a higher accuracy in the repeated trial condition for seven-letter target. Therefore, if any effect of repetition could be attributed to bias of choosing a string with no letter repetition, the accuracy of repeated letter trials should be lower only for eight-letter targets and not for seven-letter targets.

Like the study of Gomez et al., (2008), nonword stimuli were used for the intended manipulation. Unlike their study's five-letter stumili, however, the present experiment used seven- and eight-letter stimuli. The greater length was chosen for two reasons. First, longer lengths allow for variation in the positions in which the repetitions occur, thus making the items less predictable and more variable. Second, repetitions are more likely to be observed in longer words. Therefore, the processes involved in possible observed effects would be more likely to generalize to processing of longer items with repeated letters.

3.2. Method

3.2.3. Participants

The data of 96 undergraduate students from the University of Warwick were included in the analysis. All reported English as their native language. They took part in the experiment in exchange for course credit. Another 24 participants did not perform reliably above chance level (accuracy was less than 56%) and were dropped from the analyses. In addition, the data of one participant were not retrieved succesfully due to machine failure and three other participants were excess to the number that would equate counterbalancing lists. The total number of tested participants was 124.

3.2.4. Design

The major variable of interest was the presence or absence of a repetition in a trial, or *letter type* (repeated, unique). In addition, there were two different levels of *target length* (seven letters, eight letters). In each trial, the target and the foil were never the same length. The critical comparison was the one between repeated or unique letter type. When the targets were eight letters long, the critical manipulation was done in the target, which either contained repeated letters (DRARTIEN) or not (DRALTIEN). In this case, the foil was seven letters long and differed from the target in deletion of the critical letter (DRATIEN). When the targets were seven letters long (DRATIEN), and the roles of the stimuli were reversed, the repeated letters were either present or not in the foil (here DRARTIEN and DRALTIEN served as foils) and so foils differed from the target by insertion of the critical letter. Although targets and foils had different lengths, which could lead to some strategic performamce, this difference was not apparent between these two lengths, especially when nonword targets were very briefly displayed and sandwiched between a longer mask. Furthermore, this length difference was constant for both critical conditions and could not explain any possible repeated letter effects.

This design afforded testing of how participants would respond to both insertions and deletions of letters that were either already present or not in the target. The purpose of this contrast was to test the ability of the participants to detect the number of letters that shared the same identity in a brief presentation of pronounceable letter strings. To avoid the influence of possible confounds, we created 32 different lists, counterbalancing factors such as position of the correct response, occurrence of the repetition, and critical letter. These will be further explained in the next section. An illustration of all the possible conditions can be seen in Figure 8.

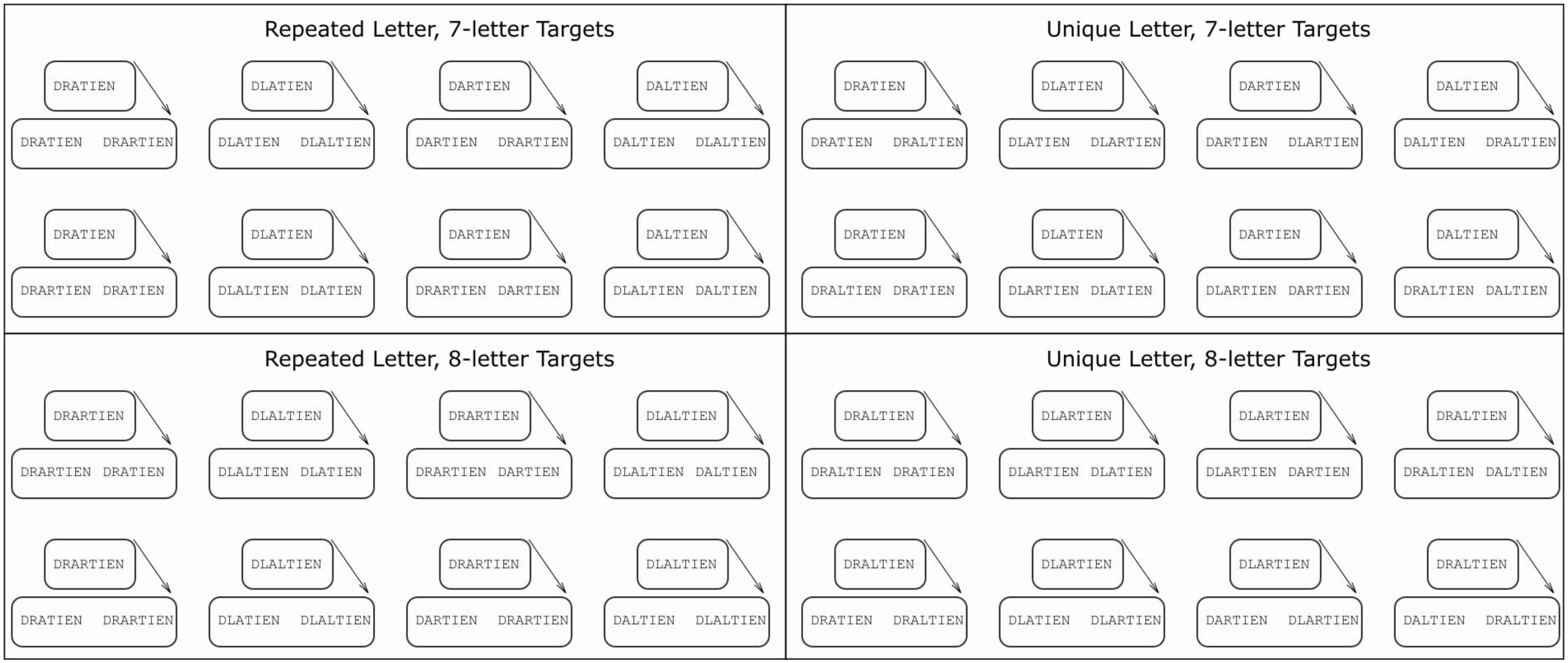

Figure 8. Conditions in Experiment 1

3.2.5. *Stimuli*

The materials[2] consisted of 352 families of 8 nonword items, which were constructed so that they were pronounceable and did not deviate substantially from the orthotactics and phonotactics of English. Each of the base pseudowords of the families was an eightletternonword that contained a repetition of a letter. The repeated letters occurred equal times (88) in positions 2 and 4; 3 and 5; 4 and 6; 5 and 7. Half of the repeated letters in each position condition were consonants and the other half were vowels. The initial and final letters were never repeated. For each of the families, 7 additional derivatives were constructed so that each family of items had 8 different members in it. The second version of each item was constructed by replacing the repeated letters with other repeated letters from the same corresponding consonant-vowel class (e.g., OLELUVAN-OBEBUVAN). The third version of the items were derived by replacing the first occurrence of the repeated letter from the first item version with the repeated letter from the second item version (e.g., OBELUVAN). The fourth item from the family was constructed in a similar way as the third one, the only difference being that the second occurrence of the repeated letter of the first version was the one that was replaced by the repeated letter of the second type (e.g., OLEBUVAN). Thus, the first four versions of the items contained two different eight-letter nonwords with repeated letters (repeated condition) and two different nonwords with no repeated letters (unique condition). All four versions were used to keep the design symmetrical as well as to counterbalance possible statistical regularities effects such as letter and bigram frequencies.

The next four versions of the item families represented one-letter-deletion derivatives of the eight-letter nonwords. They were constructed by deleting one of the letters in the critical positions where a repetition occurred in the repeated condition. The fifth version of the items were derived by deleting the first occurrence of the repeated letter of the first type (OLELUVAN-OELUVAN). The sixth item was derived by deleting the second occurrence of the first type (OLELUVAN-OLEUVAN). The seventh and the eighth versions were constructed in the same way as the previous two versions, but the first and the second occurrences of the second repeated letter type were deleted (OBEBUVAN-OEBUVAN; OBEBUVAN-OBEUVAN). The initial 352 base items were constructed so that their deletion derivatives remained pronounceable.

---

[2] Research data file including stimuli, data, analysis, DMDX scripts, simulation files of all experiments is available to download at: http://adelmanlab.org/repeated-expts/Research%20Data.zip

Each of the eight versions of an item family served as a target in a perceptual identification two-alternative forced-choice task, in which participants had to choose the correct response from a target and a foil. Each of those versions also served as a foil. There were two possible foils for each target. When the target was one of the four eight-letter targets, the two possible foils were the corresponding seven-letter versions with omission of one of the two letters in critical positions. For example, the two possible foils for OLELUVAN were OELUVAN and OLEUVAN. When the target was one of the seven-letter item versions, the two possible foils were the corresponding repeated or unique eight-letter versions. Thus, when the length of the target was seven letters, the foils had an additional inserted letter. This was the letter in the same position that was omitted to form the seven-letter target. Thus, the two possible foils for OLEUVAN were OLELUVAN and OLEBUVAN.

Each participant saw only one of the eight versions of the items as a target with one of the two possible foils. The position of the correct response and their corresponding left and right buttons was carefully counterbalanced between the lists, so that the same target-foil pair appears in both possible left-right configurations. These manipulations led to the construction of 32 different counterbalancing lists. In addition, within each list, the correct responses were equal times on the left and on the right for each of the contrasting conditions.

3.2.6. Procedure

This and the following experiments were conducted on a 17" CRT Sony Trinitron CPD-G220 monitor with the refresh rate set to 10ms. The DMDX software (Forster & Forster, 2003) was used for the presentation of the stimuli and for data collection. All stimuli were presented in black on a white background. Each trial began with a 10-symbol mask (##########), presented in the center of the screen for 900 ms in a Courier New font, size 23. The mask was followed by the target nonword, presented for 110 ms in upper Courier New font, size 20. After the presentation of the target, the screen remained blank for 10 ms after which the mask appeared again in the place of the target simultaneously with two choices which were displayed below the mask. The choices were in Courier New font, size 23 and were presented to the left and to the right of the middle of the screen. Participants had to perform a two-alternative forced-choice task. One of the alternatives was the target itself and represented the correct choice and the wrong alternative was one of the two possible foils of that target. Participants were asked to indicate which of the two alternatives was displayed on the screen by pressing either the left shift key for the alternative on the left or the right shift key for the alternative on the right. They were instructed to be as accurate as possible and had up to 2000 ms to respond. Accuracy feedback was given after each trial. In addition, participants' current percentages of correct responses were displayed after the completion of every 44 trials. Participants were encouraged to constantly try to improve their performance as much as possible and were given a break in the middle of the experiment.

3.3. Results

The accuracy results per condition can be seen in Figure 9. A generalized linear mixed-effects model with binomial distribution was fitted for the accuracy analyses of all the trials in the experiment. The model contained letter type (repeated/unique), target length (seven/eight) and their interaction as fixed effects and by-subjects and by-items intercepts and by-subjects slopes for letter type, target length and their interaction as well as by-items slopes for target length as random effects. The results revealed significant main effect of letter type, $\chi^2(1) = 34.029$, $p < .001$; significant main effect of target length $\chi^2(1) = 28.258$, $p < .001$; and a significant interaction between the two factors, $\chi^2(1) = 33.259$, $p < .001$. Participants were significantly less accurate when the targets were eight-letters long than when the targets were seven-letters long. Pairwise comparisons revealed that the repeated letter condition was significantly harder than the unique letter condition only for the trials with the shorter seven-letter targets with a letter insertion in the foils, $\chi^2(1) = 67.07$, $p < .001$. The difference between the repeated and unique letter conditions was not significant for the longer eight-letter trials with deletion in the foils, $\chi^2 < 1$.

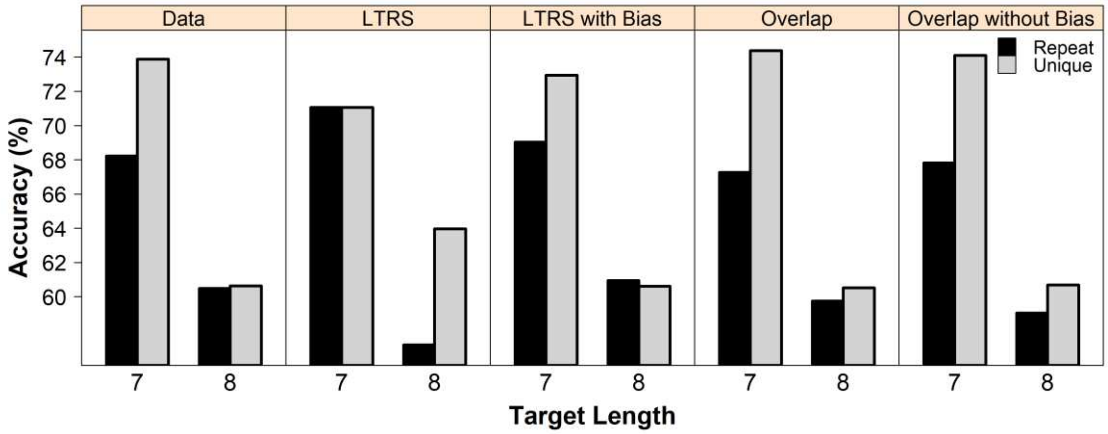

Figure 9. Percentage accuracy in Experiment 1 per letter condition (repeated vs. unique) for seven-letter targets with letter insertion in foils and for eight-letter targets with letter deletion in foils: Empirical data (from left, panel 1); LTRS prediction (panel 2); and LTRS prediction with bias (panel 3); Overlap predictions (panel 4), Overlap without bias predictions (panel 5).

3.4. Modelling

3.4.3. LTRS

Parameter-free predictions of LTRS

Insertion foils For both types of insertion foil, the discrepancy in the foil is only detectable (*v* only equals 0) when the adjacency of the letters surrounding the omission is perceived; a detailed enumeration of states is given in Appendix C's Table C1. The probability of this occurring does not depend on the identity of the foil because these correspond to the same abstract perceptual states ***x***, so these conditions are predicted to be equivalent.

Deletion foils For some of the possible perceptual states ***x*** in which the critical letter is perceived, whether or not the percept is consistent with the foil, *v*(*f*), differs by foil condition; the detailed enumeration of states is given in Appendix C's Table C2, and can be summarized as follows. In the unique case, the critical unique letter is always inconsistent with the foil, but in the repeated case, the critical repeated letter is not always distinguishable from its twin. This difference occurs when four conditions are all met: (i) the critical letter that differs between the two targets has been perceived; (ii) its twin has not been perceived; (iii) the central letter between the repetitions has not been perceived; and (iv) the adjacency of the critical letter and its non-central neighbour has not been perceived (either because the non-central neighbour has not been perceived at all or because at least one of these two letters has not reached the state of full information). There are no cases where the unique condition can produce consistency where the repeated condition does not. As such, accuracy is predicted to be higher in the unique condition.

These predictions were not borne out in the data, so we re-examined one aspect of the model.

Biasing the guess in LTRS

When considering the misfit of the original LTRS to the data of Experiment 1, we observed that our response options do not conform to a reciprocal relationship: If participants preferred to choose an option with a repeated letter, this would affect our conditions differently: Responses in the repeated-insert condition would gain accuracy at the expense of accuracy in the repeated-delete condition, while retaining the pattern of average accuracy for the two repeated conditions being greater than the average accuracy for the two unique conditions.

We added such a bias to LTRS as follows. While the probability of a correct guess g remained 0.5 on trials where neither option contained a repetition, a biased guess was made when one of the options contained a repetition, with the probability of choosing the option with the repeated letter being $0.5 + (n\text{-}r/R)b/n$ where b is a parameter controlling the size of repeated letter biases, r is the number of excess repeated letters (in the sense of: by how much the length n exceeds the number of unique letter identities) and R is the number of letter identities for which there is a repetition. In the present experiment, this was always $0.5 + 7b/8$, and this is the probability of a correct guess on repeated-deletion-foil conditions, and the probability of a correct guess on repeated-insertion-foil conditions was $0.5 - 7b/8$.

Numerical predictions of LTRS

Figure 9 includes numerical predictions of LTRS both with and without the adjustment to bias. Our initial attempts to optimize model parameters indicated that this data set did not sufficiently constrain the parameters, as some took on extreme values without substantially improving the fit of the model. Therefore, we fixed some parameters to be similar to previous fits, and constrained others to be equal. In particular, within each length, non-initial letter rates were set to be equal. The final parameters are listed in Appendix A. As can be seen in Figure 9, the original model predicted repeated letter effects only when the repetition was in the target but captured the observed data pattern once a repetition bias was implemented.

3.4.4. *Overlap Model*

An Overlap Model (Gomez et al., 2008) simulation of Experiment 1 was run after optimization of the model's parameters: the asymptote and rate parameters (controlling the standard deviations for each of the letter positions in the stimuli), scaling exponent parameters for choosing the unique letter item (a) and the repeated letter item (b). A separate simulation was also run where only one exponent was fitted (a) and so there was no implemented bias in the model for repeated letters. The parameters for both simulations are listed in Appendix B. As can be observed in Figure 9, the model captured the repeated letter effect for the shorter targets in both simulations.

Notably, even without separate parameters for the unique and repeated cases, the repetition effect for trials with eight-letter targets (that could contain a repetition) and seven-letter foils was correctly predicted to be noticeably smaller than the repetition effect with seven-letter targets and eight-letter foils (that could contain a repetition). In both cases, the overlap between the foil and the target increases when a repetition is present, because the critical letter in the longer item contributes to the overlap only if it is a repetition. As one would expect, the increased similarity between foil and target tends to increase the chances that an incorrect foil response will occur. However, the model's decision is also influenced by the overlap between the target and itself (but not the overlap between the foil and itself). Critically, unlike many other similarity schemes, self-overlaps can vary. In the foil-insertions conditions with the seven-letter targets, the target-target overlap is unaffected by the repetition that occurs in the foil – does not vary – so accuracy is just controlled by foil-target overlap, allowing for a strong repetition effect coming from the foil. In the foil-deletion conditions with the eight-letter targets, the target-target overlap is affected and is higher when the target contains a repetition. So, when the repetition deletion condition is compared with the unique deletion condition, both the foil-target overlap and the target-target overlap going into the decision rule are higher, resulting in only a mild decrease in accuracy.

In its original form with a bias affecting unique and repeated options differently, the model fitted the data slightly better, with the exponent parameter being slightly higher for repeated than unique items, which is the opposite direction of bias to the original application to Gomez et al.'s (2008) Experiment 4. Thus, the Overlap Model tended to agree with LTRS that participants were biased towards the repeated-letter options, though this was not as critical to the Overlap Model's fit as LTRS's.[3]

3.4.5. *Open bigram models and SOLAR*

The predictions of several successors of the Interactive Activation Model (McClelland & Rumelhart, 1981) were also explored by calculating match scores for the nonword pairs in the main 4 experimental conditions in Experiment 1 (these also apply to the main contrasts in Experiment 2 and Experiment 3). The match scores calculations were obtained with Davis's Match Score calculator (http://www.pc.rhul.ac.uk/staff/c.davis/utilities/matchcalc/index.htm) for the SOLAR model (Davis, 1999), as well as for two versions of the overlap bigram scheme: open bigram (Grainger & van Heuven, 2003), and the overlap open bigram (Grainger et al., 2006). As can be seen in Table 5, SOLAR predicts no difference between the repeated and unique conditions. The open bigram models, however, predict stronger similarities in the cases including repeated letters than the cases with no repetition in both target length conditions. In addition, all models generally predict stronger similarity between the two strings when the comparison word is the shorter stimulus and the longer stimulus is compared to the shorter one, than in the opposite case, suggesting an asymmetric relationship between the strings in the pair.

Table 5. Mean match score calculations by target length and letter type for SOLAR (Davis, 1999), Open Bigram (Grainger & van Heuven, 2003), Overlap Open Bigram (Grainger et al., 2006)

Target Length	**7**		**8**	
Model	Repeated DRATIEN-DRARTIEN	Unique DRATIEN-DRALTIEN	Repeated DRARTIEN-DRATIEN	Unique DRALTIEN-DRATIEN
SOLAR	0.97	0.97	0.873	0.873
Open Bigram	0.9	0.85	0.845	0.71
Overlap Open Bigram	0.978	0.883	0.78	0.753

[3] Conversely, when we fitted the new version of LTRS with bias to Gomez et al.'s (2008) Experiments, LTRS agreed that there was a slight bias against the repeated-letter items. The improvement in fit compared to the original LTRS was, however, fairly small.

3.5. Discussion

The results of Experiment 1 showed that participants performed significantly worse when an additional inserted letter in the foil was already present in the target, therefore producing a letter repetition, then when the inserted letter was different from any letters in the target. These results suggest that letter numerosity might be difficult to process at early stages of orthographic processing. A bias toward choosing a string with no repeated letters was not observed in these conditions. Such a bias was offered as an explanation by Gomez et al. (2008) for the observed lower accuracy for targets with repeated letters and foils with no repeated letters in a two-alternative forced-choice perceptual identification task. Two separate simulations with the Overlap Model showed that the results of Experiment 1 could be captured by the model with and without an implemented bias. The results of Experiment 1 also indicated lower accuracy for the eight-letterg targets than for the seven-letter targets, suggesting that the longer targets were harder to perceive. In the trials, in which the foil was missing a letter from the target, the accuracy was not influenced by the type of the missing letter, repeated or unique. The difference in the pattern of the repeated letter effect in the two target length conditions signals for orthographic processing asymmetries between insertions and deletions string pairs. LTRS (Adelman, 2011) predicted a difference between the two letter type conditions only when the critical letter was present in the target. This prediction represented the opposite pattern of the observed data and was based on the original implementation of the model for forced-choice perceptual identification task with two alternatives. The presence of the repeated letters effect only for the shorter targets and the inconsistency with the predictions of LTRS could be explained by generally higher processing difficulty for items with repeated letters that could in turn cause a strategic preference for the repeated item as a form of overcompensation. This hypothesis was tested in a subsequent simulation with a version of LTRS augmented with this bias. This model captured the observed data pattern with the inclusion of a slight preference for the repeated letter item when the model guessed because the percept was ambiguous. The susceptibility to bias-based explanations is a feature of the task (not peculiar to LTRS) and shows that this task is not ideal for comparisons where contrast of interest is not symmetrical between response options.

4. Experiment 2: Same-Different Task with Deletion Primes

The next two experiments followed up on the evidence from Experiment 1 that distingushing repeated letter identities is perceptually more demanding than distinguishing corresponding different letter identities. The results of Experiment 1 demonstrated effects of repeated letters with a two-alternative forced-choice perceptual identification task with a manipulation of letter insertion in the foil. This result suggested that the two choices were perceived as more similar in the repeated letter condition than in the unique letter condition. As already discussed, it is possible that the different results with deletions and insertions in the foils could be attributed to adjusted strategic performance that compensates for processing limitatation in cases of repeated letters.

Experiments 2 and 3 aimed to further explore the repeated letter effects with deletion and insertions with a paradigm that is less susceptible to conscious strategies. They employed the masked priming paradigm combined with the same-different task. We used the same-different task for several reasons. First, it could be combined with the masked-priming paradigm and provide priming results that are not affected by strategic effects and biases. Furthermore, unlike the lexical decision task in which the priming effects are usually restricted to word targets, robust masked form priming effects occur with the same-different task for both word and nonword targets (Kinoshita & Norris, 2009). This ensures the compatibility of the task with the nonword stimuli used in Experiment 1 and affords for a systematic exploration of the effect across the different paradigms. Finally, it has been demonstrated that the masked-priming same-different task is sensitive to small orthographic manipulations, such as effects of repetition, but these effects were demonstrated only with adjacent conditions (Norris et al., 2010). The focus of the present work was to establish whether there was a more general mechanism involved in a differential processing between two nonadjacent repeated identities and two different identities in a letter string. The aim was to establish whether the previous repeated letter findings observed with the masked-priming same-different task could be extended to nonadjacent repetitions in the context of the insertion and deletion manipulations used in the perceptual identification task in Experiment 1.

The purpose of Experiment 2 was to test whether a repeated letter effect, analogous to the one observed in Experiment 1, could also be demonstrated with the masked-priming paradigm and a deletion manipulation in the prime. Experiment 2 explored whether there was a higher orthographic similarity between a deletion prime and target with repeated letters than between a deletion prime and target with no repeated letters. The effect of identity primes on the two target types was also tested to make sure that any deletion priming differences could not be attributed to one of the target types being more prone to priming in general.

4.2. Method

4.2.3. Participants

Sixty-four native English speakers took part in the experiment for a small payment and were included in the analyses. Two other participants were excluded due to low accuracy scores (correct on less than 70% of the trials).

4.2.4. Design

As the task was to determine whether two letter strings, a reference and a target, were same or different, there were two levels of trial types that occurred equal times: same and different. In addition, the design of the experiment contained three more factors: 2 x *target types* (repeated letters, unique letters), x 2 *prime relatedness* (related, control) x 2 *prime length* (7 letters, 8 letters) x 2 trial type (same, different). In addition, the different trials had two types of *reference type* (repeated, unique). The four different prime types comprised of related identity and deletion primes and their corresponding control unrelated conditions. Each participant saw only one version of an item in only one of the possible conditions but was presented with all the different conditions in the design. There were 32 different counterbalancing lists. Examples of the conditions in same trials could be seen in Figure 10.

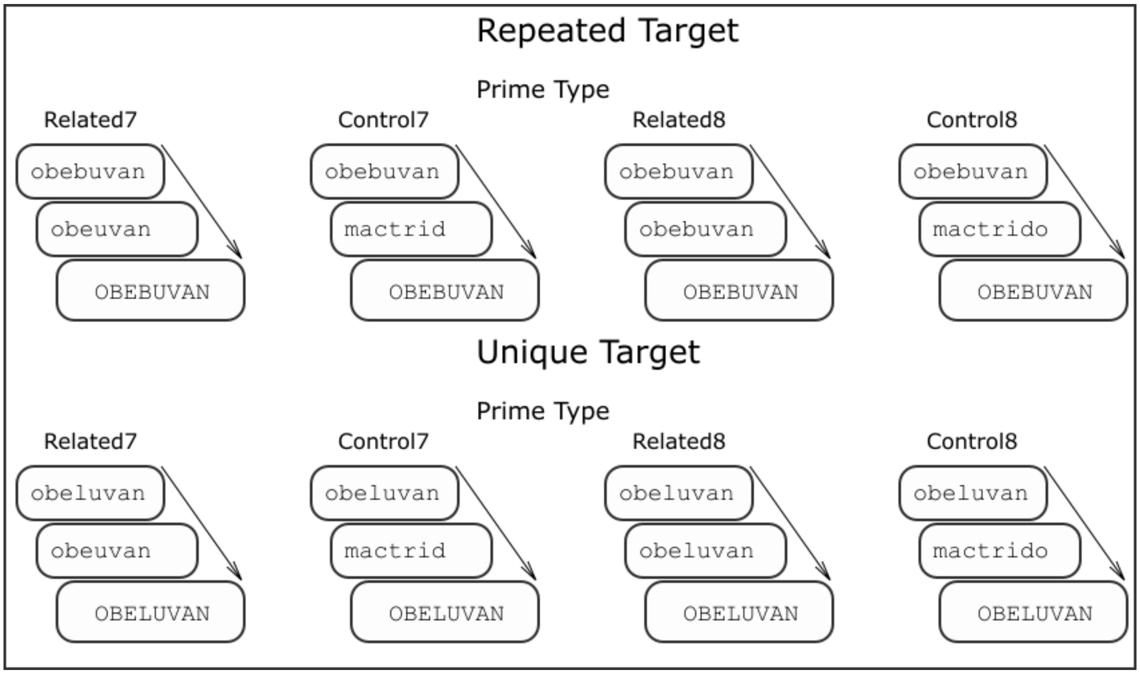

Figure 10. Conditions in Experiment 2, Same trials.

4.2.5. Stimuli

Stimuli were re-used from Experiment 1. To minimize the counterbalancing list conditions, from each item family only one of the two eight-letter-long versions containing repetition and only one not containing repetition were selected as targets for both the same and different trials. They also served as references in the same trials. In the same trials, the eight-letter related prime was identical to the target and the reference. The seven-letter related prime was the same as the identity with omission of a single letter. The omitted letter was the different one between the repeated letter family version (e.g. OBEBUVAN) and the nonrepeated letter one (e.g. OBELUVAN), so the relationship between the seven-letter related prime (e.g. obeuvan) and the two target types was the same: The prime contained seven out of eight letters of the target. The control primes were constructed by pairing each family with another family so that both families shared no more than three common letters. Where it was not possible to pair families so that control primes shared two or fewer letters with their target, one letter was changed in the prime to meet this constraint. After those manipulations, the control primes contained no more than two common letters with the targets. The seven-letter control primes were constructed by deleting one common letter from the eight-letter unrelated primes. Where there was a choice, the letter whose omission preserved pronounceability was chosen. The seven-letter unrelated primes had no more than one common letter with the target.

The previously unselected eight-letter versions of an item family served as references for the different trials. There were two different reference types. The reference either contained a repetition for which the second repeated letter version from a family was selected (e.g., abebulit from the ARERULIT family) or not, for which the second unique letter version from the family was selected (e.g., olebuvan from the OLELUVAN family). The two different reference types occurred equal times in the different trials for each target type, so the outcome could not be determined only be the presence or absence of a repetition in the reference. For half of the items the repeated letter reference was selected and for the other half the unique letter reference was selected. A zero-contingency scenario was adopted for the different trials. In this scenario, the correct response cannot be predicted by the relationship between the reference and the prime, as in both trial type conditions, the prime is related to the reference. In the different trials, the identity primes were the same as the reference, the seven-letter related prime was more related to the reference than to the target (e.g., apoplecy-apolecy-ARORLECY) and the two control prime conditions were the same as the control primes in the same trials and were neither related to the reference nor to the target.

4.2.6. Procedure

Each trial began with a presentation of a 10-symbol (##########)mask in the center of the screen and the reference above it. These were presented for 1 s after which the reference disappeared, and the prime replaced the mask. The prime was displayed for 60 ms and it was followed by the target. The target stayed on the screen until response with a 2-s timeout. The reference and the prime were presented in lowercase and the target was presented in uppercase. Courier New font was used, sizes 12.5 for the reference and the prime, and 20 for the target. Participants were instructed to respond as quickly and as accurately as possible whether the pair of nonwords they saw on the screen were the same or different by pressing one of two corresponding buttons (left shift for different, right shift for same). They were instructed to disregard the difference in the case. The presence of the primes was not mentioned.

4.3. Results

4.3.3. Same Trials

Response Time. Prior to the response time analyses, trials with incorrect responses (8.7%) and response times faster than 150 ms and slower than 1500 ms were removed (1.3% of the correct trials). Mean response times and error rates by condition are displayed in Table 6. A linear mixed-effects model was fitted with target type (unique/repeated), prime relatedness (related/control), prime length (7 letters/ 8 letters) and their interaction as fixed effects and by-subjects and by-items intercepts as random factors. The random slopes were excluded from the model as it failed to converge.

Table 6. Mean Response Times (ms), Error Rates (%) by Condition and LTRS priming predictions for Experiment 2, Same Trials

	Prime Length			
	7		**8**	
	Deletion/Control		**Identity/Control**	
Target Type	Unique	Repeated	Unique	Repeated
Prime				
Related	735(8.0)	742(8.2)	739(8.3)	743(7.9)
Control	763(9.9)	767(8.8)	760(9.3)	764(8.9)
Priming	28(1.9)	25(0.6)	21(1)	21(1)
LTRS	51	51	63	63

The effect of prime relatedness was significant, $\chi^2(1) = 35.153$, $p < .001$. The effect of target type was not significant $\chi^2(1) = 1.859$, $p = .173$. The effect of prime length and all the interactions were not significant, $\chi^2 < 1$. These results suggested that participants were not significantly delayed in the repeated letters target condition. They were also equally primed by the identity 8-letter primes and by a deletion 7-letter primes and the priming effect did not differ across target type conditions.

Accuracy. For the accuracy analyses, a generalized linear mixed-effects model with binomial distribution was fitted with the same structure as the model for the response time analyses. The results revealed a main effect of prime relatedness, $\chi^2(1) = 4.89$, $p = .027$. Participants produced significantly more errors when the primes were unrelated than when the primes were related. No other results were significant, all $\chi^2 < 1$.

4.3.4. *Different Trials*

Response Time. Prior to the response time analyses, trials with incorrect responses (20.2%) and response times faster than 150 ms and slower than 1500 ms were removed (1.7% of the correct trials). Response times for the different reference and target type conditions are displayed in Table 7[4]. A linear mixed-effects model was fitted with target type (unique/repeated), prime relatedness (related/control) and prime length (7 letters/ 8 letters), reference type (repeated/ unique) and their interaction as fixed effects and by-subjects and by-items intercepts as random factors. The results revealed a significant main effect of target type, $\chi^2(1) = 510.344$, $p < .001$; reference type, $\chi^2(1) = 486.748$, $p < .001$; and significant interaction between target type and reference type, $\chi^2(1) = 17.869$, $p < .001$. The effect of prime length approached significance, $\chi^2(1) = 3.01$, $p < .083$.

As can be seen in Table 7, participants were fastest in the condition in which both the reference and the target contained a repetition. In this condition, the reference was two-letter different from the target, as the repeated letter identity was replaced with a different one (dlaltien-DRARTIEN). Separate contrasts between this condition and the conditions with one letter difference between the reference and the target confirmed a significant delay, $\chi^2(1) = 129.43$, $p < .001$; $\chi^2(1) = 178.86$, $p < .001$; for contrasts with the repeated-unique (dlaltien-DLARTIEN) and unique-repeated (olebuvan-OBEBUVAN) reference type-target type conditions, respectedly. Crucially, the difference between the latter two conditions was also significantly different, $\chi^2(1) = 7.52$, $p = .006$. This result suggested that when there was a letter repetition only either in the reference or in the target, the repetition is less resource-demanding if its representation has been established (presented as the reference) rather than when the repeated letter string does not have an established representation and is perceived for the first time (presented as a nonword target in a different trial). The remaining pairwise contrasts confirmed

[4] A more detailed table for the different trials in Experiment 2 is included in the supplementary material.

that the most difficult reference type-target type condition, which contained letter transposition and no repeated letters (olebuvan-OBELUVAN) was significantly slower than the other three conditions, $\chi^2(1) = 305.27; 52.955; 24.098$, all $p < .001$.

Table 7. Response Times (ms) and Error Rates (%; in parentheses) by Reference and Target Type in Different Trials, Experiment 2

Reference	Target	RT/Error Rate mean	RT prime 7	RT prime 8
Repeated (apoplecy)	Unique (AROPLECY)	775 (20)	783	766
Unique (olebuvan)	Repeated (OBEBUVAN)	808 (22)	812	805
Unique (olebuvan)	Unique (OBELUVAN)	877 (34)	874	879
Repeated (apoplecy)	Repeated (ARORLECY)	686 (06)	688	684

Accuracy. For the accuracy analyses, a generalized linear mixed-effects model with binomial distribution was fitted with target type, reference type and their interaction as fixed effects and by-subjects and by-item intercepts as random effects. The prime relatedness and length were excluded from the model as it failed to converge even after dropping the random slopes. The results revealed a main effect of target type, $\chi^2(1) = 304.852$, $p < .001$, a main effect of reference type, $\chi^2(1) = 211.578$, $p < .001$, and a significant interaction between the two factors, $\chi^2(1) = 57.099$, $p < .001$. A pairwise contrast between the repeated-unique and unique-repeated reference type-target type conditions revealed that the 2% difference was significant, $\chi^2(1) = 4.749$, $p = .029$. Accuracy was significantly lower when the repetition was in the target and not in the reference than vice versa.

4.4. Modelling

4.4.3. LTRS

Parameter-free predictions of LTRS

For Experiment 2, the analysis for deletion primes is the same as that for the insertion-foil conditions of Experiment 1 (when the briefly presented stimulus is 7 letters long; Table C1 of

Appendix C) with the briefly presented prime in Experiment 2 taking the role of the briefly presented target in Experiment 1.

That is, the two different targets are not affected differently because the inconsistency is in the absence of a letter. This is detected when the adjacency of the two surrounding letters is detected, regardless of the identity of the letter in the target. Therefore, priming is predicted to be identical for repeated and unique conditions, and this is a categorical prediction of the model, not dependent on parameters.

Likewise, identity priming cannot differ for targets of the same length, as no discrepancies can occur (but β parameters can depend on the length).

Numerical predictions of LTRS

It would be possible to choose LTRS parameters that produce any positive amount of priming from seven-letter deletion primes (e.g., 26ms), and any positive amount of identity priming for eight-letter stimuli (e.g., 21ms) if no restriction were placed on the β parameters.

Instead, for purposes of illustration, Table 6 includes the numerical LTRS priming predictions that come from the parameters that Adelman (2011) optimized for word priming in the lexical decision task. The model predicted identical priming effects for the repeated and unique conditions. This prediction captured the observed lack of difference between the repeated and unique conditions. The predicted bigger priming effect for the identity eight-letter primes than for the deletion seven-letter primes, however, was not observed in the data.

4.4.4. Overlap Model

Overlap scores between the targets and each of the prime types were calculated using the asymptote and rate parameters suggested by Gomez (2020) to control the standard deviations. As can be seen in Table 8, the identity value was higher for a repeated letter target, suggesting that this target type overlaps more with itself than the unique letter target type, due to the repeated letter. The related deletion prime also had a higher overlap scores for the repeated letter targets than for the unique letter targets. As the values for the identity values were not the same for both conditions, and they were not fixed to 1, the interpretation of the predictions was not straightforward. These predictions, however seemed at odds with the observed lack of difference between the repeated and unique cases in both the identity and the deletion prime conditions.

Table 8. Overlap scores for Experiment 2

Target	Identity 8-letter	Control 8-letter	Related 7-letter	Control 7-letter
Repeated	2.86	0.21	2.23	0.1
Unique	2.66	0.21	2.08	0.1

4.5. Discussion

The results of Experiment 2 showed similar deletion and identity priming effects for the repeated and unique target types. These results did not appear to be consistent with the predictions of the Overlap Model (Gomez et al., 2008) and also disagreed with the higher match scores for the repeated conditions than the unique given by open bigram schemes. The lack of difference between the repeated and unique conditions was predicted by LTRS and was consistent with previous studies that have investigated the same effect with such subset primes in the lexical decision task (Schoonbaert & Grainger, 2004). Furthermore, there was no effect of target type in the response time analysis in the same trials, suggesting that both types of targets took the same time to process, therefore showing no effect of repeated letters. The different trials, however, provided some important results regarding repeated letters. The evidence suggests that the repeated letters are more difficult to process when the string in which they are embedded does not have an established representation than when the presence of a repetition is anticipated.

5. Experiment 3: Same-Different Task with Insertion Primes

Experiment 3 tested whether a repeated letter effect could be established if the relationship between the prime and the target was the reversed case of the one in Experiment 2. The longer stimulus with the additional letter was the prime (the brief event), and the shorter stimulus served as the target. Experiment 3 explored whether a possible masked priming effect with repeated letters was asymmetrical, and whether it could be affected by the anticipation of the repetition as well as its processing time. We tested whether the orthographic similarity between insertion primes and targets could be affected by the status of the inserted letter (repeated vs unique). This experiment differs from previous studies with similar manipulation by the task (same-different rather than lexical decision) and the lexicality of the targets (nonwords rather than words). The aim was to establish whether the insertion prime with repeated letters would produce stronger priming effect than insertion primes with no repeated letters. This expectation was generated by the results of Experiment 1, suggesting that foils with an inserted letter already present in the target were more similar to targets than foils with an inserted letter that was not present in the target. It was also caused by the results of the different trials in Experiment 2, suggesting that repeated letters might be more difficult to process when their presence is not anticipated. As in the previous two experiments, the repetitions of letters in the insertion primes were nonadjacent with one intervening letter between the two repeated ones.

5.2. Method

5.2.3. Participants

Seventy-five native English speakers took part in the experiment for a small payment. The last three participants were added to complete the counterbalancing after three were dropped from the sample due to low accuracy scores (correct on less than 75% of the trials), leaving data from 72 for analysis.

5.2.4. Design and Stimuli

The same set of items as those in Experiment 2 were used. However, this time the seven-letter items served as targets and references and were primed by the eight-letter items. The references and the primes were identical for the same and different trial types. There were three different prime types: A related prime, containing an insertion of a letter already present in the target (e.g., obebuvan-OBEUVAN); a related prime, containing an insertion of letter not already present in the target (e.g., obeluvan-OBEUVAN); a control prime (e.g., mactrido-OBEUVAN). For the different trials, the alternative seven-letter version was chosen from the family set as a target. It differed by the same trial target and the reference by only one letter (e.g., obeuvan-mactrido-OLEUVAN). Examples of the conditions in same trials could be seen in Figure 11. The different trials differed from the same trials only by the target (e.g., OLEUVAN).

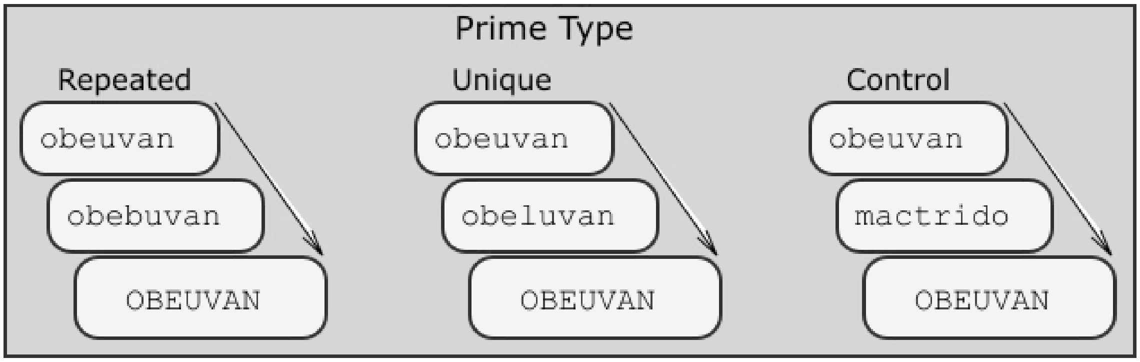

Figure 11. Conditions in Experiment 3, Same trials.

5.2.5. Procedure

The procedure was identical to the one in Experiment 2.

5.3. Results

5.3.3. Same Trials

Response Time. Prior to the response time analyses, trials with response times faster than 150 ms and slower than 1500 ms (0.47%) and incorrect responses (7.43%) were removed. Mean response times and error rates by condition are displayed in Table 9. A linear mixed-effects model was fitted with a prime type (unrelated/ related repeat/ related unique) as a fixed effect and by-subjects and by-items intercepts and slopes for prime type as random factors (the full model). The effect of prime type was significant, $\chi^2(2) = 83.402$, $p < .001$. A post-hoc pairwise comparison between the three conditions revealed that the difference between the repeated letter related prime conditions and the unrelated prime condition was significant, $\chi^2(1) = 80.001$, $p < .001$, as was the difference between the unique letter prime and the unrelated prime, $\chi^2(1) = 47.829$, $p < .001$. The difference between the two related primes was also significant, $\chi^2(1) = 7.52$, $p = .006$, with participants responding significantly faster in the related repeat condition than in the related unique condition.

Table 9. Mean Response Times (ms) and LTRS priming predictions for Experiment 3, Same trials

	Response Times	Priming (ms)	LTRS
Prime Type			
Control	656		
Related repeat	618	38	32
Related unique	627	28	27

Accuracy. The error rates for the control, related repeat and related unique conditions were 9%, 7.1% and 6.6%, respectively. For the accuracy analyses, a generalized linear mixed-effects model with binomial distribution was fitted with prime type as a fixed effect and by-subjects and by-items intercepts and slopes for prime type as random effects (the full model). The effect of prime type was significant, $\chi^2(2) = 9.621$, $p = .008$. Post-hoc tests revealed that the difference between the repeated letter related prime and the unrelated prime was significantly different, $\chi^2(1) = 7.847$, $p = .005$, as was the difference between the related unique prime and the unrelated prime, $\chi^2(1) = 6.292$, $p = .012$, but not the difference between the two related prime type conditions ($\chi^2 < 1$). People produced significantly fewer errors in the two related prime conditions than in the unrelated prime condition.

5.3.4. Different Trials

Response Time. Prior to the response time analyses, trials with response times faster than 150 ms and slower than 1500 ms (0.70%) and incorrect responses (11.21%) were removed. The mean response times in the different trials for the control, related repeat and related unique conditions were 676 ms, 682 ms, and 679 ms, respectively. A linear mixed-effects model was fitted with prime type (unrelated/ related repeat/ related unique) as a fixed effect and by-subjects and by-items intercepts and by-items slopes for prime type as random factors (the by-subjects slope for prime type was dropped due to converge failure). The effect of prime type was not significant, $\chi^2(2) = 2.259$, $p = .323$.

Accuracy. In the different trials, the error rates for the control, related repeat and related unique conditions were 9.9%, 12.8% and 11.7%, respectively. For the accuracy analyses, a generalized linear mixed-effects model with binomial distribution was fitted with prime type as a fixed effect and by-subjects and by-items intercepts and slopes for prime type as random effects (full model). The effect of prime type was significant, $\chi^2(2) = 12.896$, $p = .002$. Post-hoc pairwise contrasts revealed significant difference between the repeated letter related prime and the unrelated prime, $\chi^2(1) = 12.879$, $p < .001$, and between the unique related and unrelated prime, $\chi^2(1) = 4.213$, $p = .040$. The difference between the two related conditions did not reach significance, $\chi^2(1) = 2.819$, $p = .093$. People produced significantly more errors in the related prime conditions than in the unrelated prime conditions.

5.4. Modelling

5.3.1. LTRS

Parameter-free predictions of LTRS

For Experiment 3, the analysis for insertion primes is the same as that for the deletion-foil conditions of Experiment 1 (when the briefly presented stimulus is 8 letters long; Table C2 of Appendix C) with the briefly presented prime in Experiment 3 taking the role of the briefly presented target in Experiment 1.

That is, in some of the cases where the unique inserted letter is perceived and causes an inconsistency, the corresponding case for the repeated unique letter is ambiguous (as it could be the other instance of the same identity; These are described in §3.3.1) and does not cause an inconsistency. There are no cases where the repeated insertion causes an inconsistency and the unique insertion does not. Therefore, priming is predicted to be greater for the repeated condition than the unique condition, and this is a categorical prediction of the model, not dependent on parameters.

Numerical predictions of LTRS

It would be possible to choose LTRS parameters that produce any positive priming predictions where the repeated insertion produces greater priming than the unique insertion (i.e., fit the priming summary in Table 9 perfectly).

Instead, for purposes of illustration, Table 9 includes the numerical LTRS priming predictions that come from the parameters that Adelman (2011) optimized for word priming in the lexical decision task. Unlike Experiment 2, the model's predictions for Experiment 3 was bigger priming effect for the repeated condition than unique conditions. This prediction matched the observed data qualitatively. LTRS predicted a smaller effect in the case of the unique letter insertion as the cessation of priming in this case was affected by perceiving one wrong letter identity (the inserted letter), while in the case of the repeated letter insertion, the detection of the inserted letter with consistent identity as well as additional contextual information is required to terminate the accumulation of priming.

5.3.2. Overlap Model

As for Experiment 2, we evaluated the predictions of the Overlap Model (Gomez et al., 2008) by calculating overlap scores between the target and the primes in Experiment 3. Again, we used the asymptote and rate parameters suggested by Gomez (2020) to set the standard deviations for the model. As can be seen in Table 10, the overlap was higher between the repeated letter insertion prime and the target than between the unique letter insertion prime and the target.

Table 10. Overlap scores for Experiment 3

	Repeated 8-letter	**Unique 8-letter**	**Control 8-letter**
Overlap score	2.77	2.60	0.24

5.5. Discussion

The most important finding in Experiment 3 was the stronger priming effect produced in the repeated letter condition than in the unique letter condition. To the best of our knowledge, this is the first reported case of nonadjacent repeated letter effects obtained with the masked priming paradigm. The results showed that a repeated letter effect could be obtained with an insertion manipulation in the prime when the repetition was not anticipated and had a limited processing time. The effect was qualitatively captured in the predictions of LTRS (Adelman, 2011). It was also in accordance with the predictions of the Overlap Model (Gomez et al., 2008) and as can be seen in Table 5, the results were in line with the match scores of open bigram models (Grainger et al., 2006; Grainger, van Heuven, 2003), but not with SOLAR (Davis, 1999).

6. Experiment 4: Perceptual Identification in a Same-Different Task

The following experiment continued investigating the repeated letter effects with a task that was a hybrid between the previous two tasks. As in the case of Experiment 1, the task in Experiment 4 was also perceptual identification. However, the presentation of the target was followed by only one option, which was either the same as the target or a different one, thus resembling the same-different task in Experiment 2 and Experiment 3. The perceptual identification same-different task was selected for two reasons. First, to minimize the role of decision processes that involve two alternatives. Second, to minimize top-down effects which might be present in a task with a reference stimulus and might influence early perceptual processes. The aim of this experiment was to further explore processes underlying the encoding of letter position and letter identity with a different paradigm and to provide further information of how processing difficulty might differ depending on whether a string contains repeated letter identities or not. This time, the contrasted conditions always included stimuli of the same length, rather than of pairs with missing or additional letter, as in the previous experiments. The purpose was to test whether the visual system discriminates equally well between two items, target and foil, depending on whether the target had repeated letters or not, in two different new scenarios with two new foil types. The observed data was compared to the predictions of LTRS (Adelman, 2011) and the Overlap Model (Gomez et al, 2008), as well as to match scores from open bigram (Grainger et al., 2006; Grainger, van Heuven, 2003) and spatial coding schemes (SOLAR; Davis, 1999).

6.2. Method

6.2.3. Participants

The analyses included the data from 96 participants. They were all native English speakers who took part in the experiment for course credit. The data from another 24 participants were excluded as they did not perform reliably above chance level (correct on less than 56% of the trials). Those participants were replaced to equalize the observations per counterbalancing list. The total number of participants was 120.

6.2.4. Design

Due to the nature of the task, the design included two different *trial types*: same and different. In the different trials, there were two *foil types*: substitution and wrong repetition. In the first case, the foil contained two substituted letter identities with preserved positional and identity information in the nonsubstituted letters (unique target-foil pair: abcdefgh-ab**ij**efgh; repeated target-foil pair: adcdefgh-ad**ij**efgh). In the second case, there were no new inconsistent letter identities in the foil. However, there were positional and numerosity violations resulting from deletion of one letter and insertion of a different consistent one (unique: abcdefgh-a**c**bcefgh; repeated: adcdefgh-a**c**dcefgh). In addition, the design included two *presentation type* conditions: normal and enhanced. In the enhanced condition, the critical letter was presented briefly (20 ms) in its position before the presentation of the whole target string. The purpose was to test whether any possible resolution of temporal perceptual ambiguity could interact with the type of the critical letter (unique or repeated). Examples of the conditions could be seen in Figure 12 for same trials and Figure 13 for different trials.

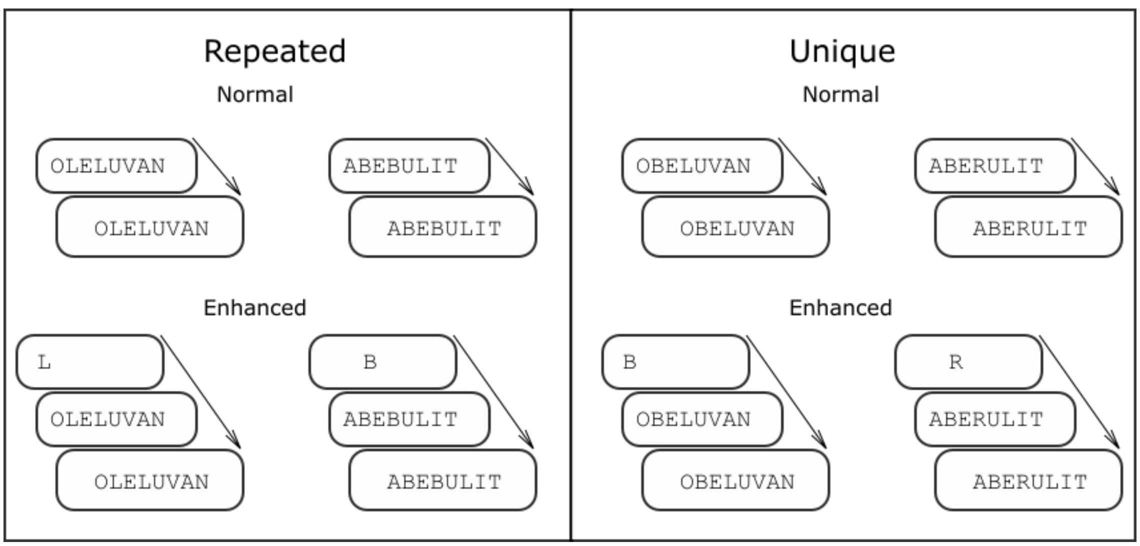

Figure 12. Conditions in Experiment 4, Same trials. Examples with two different items.

6.2.5. Stimuli

A subset of 256 family items from the materials of Experiment 1-3 were used to form two different target types: targets with repeated letters (OLELUVAN) and targets with unique letters only (OBELUVAN). All targets had eight letters. As in the previous experiments, the repeated letters in the repeated target condition were always separated by one unique letter. In half of the items, the repeated letter was a consonant and in the other half it was a vowel. In addition, each of the consonant-vowel half was further divided into two groups, so that in half of those items the unique letter target was formed by changing the first occurrence of the repeated letter (as in OLELUVAN-OBELUVAN), while in the other half the change was in the second occurrence (ABEBULIT-ABERULIT). In each of the consonant-vowel and occurrence subgroups, the repeated letters appeared equal times in critical position pairs (one replaced and one nonreplaced) 2 and 4; 3 and 5; 4 and 6; 5 and 7.

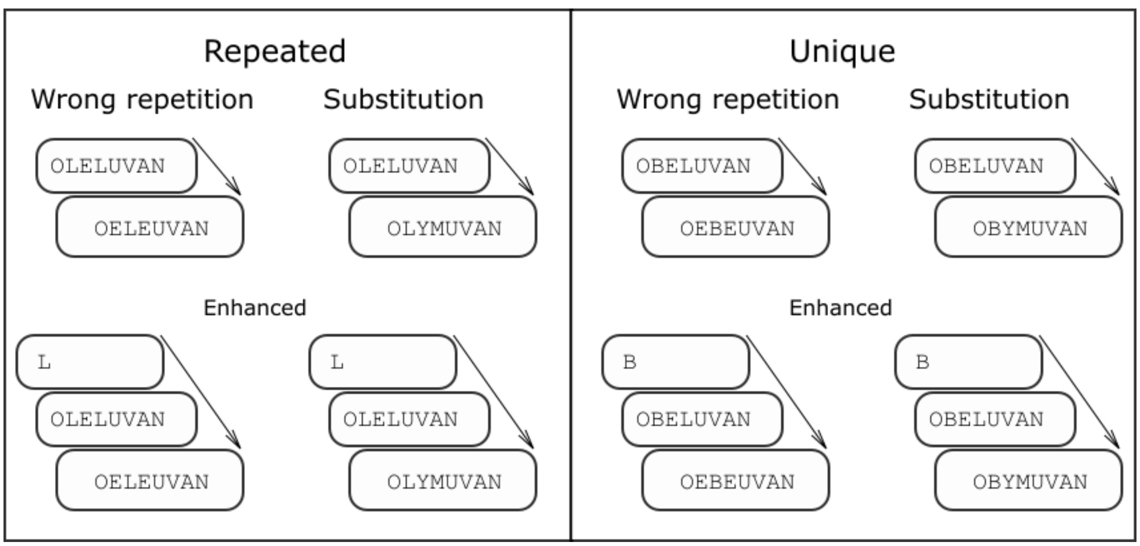

Figure 13. Conditions in Experiment 4, Different trials.

As the task required a comparison between two strings (one-alternative-forced choice; are they same or different?), different alternatives (foils) were also created for the different trial condition. These were also two different types, foils with repeated letters (wrong repetition foils) and foils with unique letters only (substitution foils). The wrong repetition foil was created by first deleting the repeated or unique letter in the critical replacement position (OLELUVAN and OBELUVAN will become OELUVAN) and then inserting a copy of the intervening letter in the other critical position. Thus, the wrong repetition foils for the targets

OLELUVAN and OBELUVAN are: O**ELE**UVAN and O**EBE**UVAN, respectively (see Figure 13). The items with a second occurrence replacement, such as ABEBULIT-ABERULIT will form A**EBE**ULIT and A**ERE**ULIT as wrong repetition foils. The substitution foil was formed by substituting two of the target letters with other two unique letters. The substituted letters were always the intervening letter (between the two critical positions) and the unchanged occurrence of the repeated letter that is common between the repeated letter type target and unique letter type target (e.g. the L in position 4 in OLELUVAN-OBELUVAN). The corresponding unique letter substitution foils for the repeated letter target OLELUVAN is OL**YM**UVAN and for the unique letter target OBELUVAN the substitution foil is OB**YM**UVAN.

In the enhanced presentation condition, the letter in the critical replaced position was briefly presented before the target. Thus, in the repeated letter target condition, the presentation of the target OLELUVAN was preceded by the short presentation of letter L in position 2, while in the unique letter target condition, the target OBELUVAN was preceded by the short presentation of letter B in position 2. The enhanced letters were identical for the different trial condition.

6.2.6. Procedure

The procedure of Experiment 4 resembled the one of Experiment 1 with a few differences that reflected the number of the alternatives (one, instead of two), the enhancement conditions, and a slightly longer presentation of the targets (120 ms, instead of 110 ms). The target duration was longer due to the increased difficulty in Experiment 4, in which all the targets had eight letters. The same apparatus and software were used as in the previous experiments. Each trial began with a 10-symbol mask (##########), presented in the center of the screen for 900 ms in a Courier New font, size 23. In the normal presentation condition, the mask was followed by the target nonword, presented for 120 ms in upper Courier New font, size 20. In the enhanced presentation condition, the presentation of the target was preceded by 20 ms presentation of the enhanced critical letter. After the presentation of the target, the screen remained blank for 10 ms after which the mask appeared again in the place of the target simultaneously with one comparator which was displayed below the mask. The comparator was in Courier New font, size 23. Participants had to perform a forced-choice same-different task. They had to decide whether the target and the comparator were the same or different and were asked to respond by pressing one of two corresponding keys. In half of the counterbalancing lists, the left shift was the key for different and the right shift was the key for same, and in the other half the keys were switched[5]. They were instructed to be as accurate as possible and had up to 2000 ms to respond. Accuracy feedback was given after each trial. Participants were encouraged to constantly try to improve their performance as much as possible and were given a break in the middle of the experiment.

6.3. Results

6.3.3. Same Trials

The accuracy results per condition for same trials can be seen in Table 11. A generalized linear mixed-effects model with binomial distribution was fitted for the accuracy analyses with target type (repeated/unique), presentation (enhanced, normal) and key allocation (right-same or left-same) and all their interactions as fixed effects and by-subjects and by-items intercepts and slopes for target type (repeated/unique), presentation and their interaction as random effects. The results revealed significant effects of target type, $\chi^2(1) = 8.347$, $p = .004$, and presentation, $\chi^2(1) = 4.845$, $p = .028$. No other results were significant, $\chi^2(1) < 1$. Overall, in the same trials, participants were significantly more accurate when the target contained repeated letters. They were also significantly facilitated by the presentation of the single letter enhancement.

Table 11. Mean Accuracy (%) by Condition in Same Trials, Experiment 4

	Presentation	
Target type	Normal	Enhanced
Repeated	78.2	79.4
Unique	75.7	77.4

6.3.4. Different Trials

[5] Data were first collected with counterbalancing lists in which the left key was different, and the right key was same. Subsequently, the keys were counterbalanced to make sure that the data pattern was not affected by key allocation through effects such as differential guessing for the two positions. As the pattern remained broadly the same, here we report the combined results with key allocation included as a fixed effect. Tables with results from the separate experiments (4a and 4b) are included in the supplementary material.

The accuracy results per condition for different trials can be seen in Table 12. A generalized linear mixed-effects model with binomial distribution was fitted with target type (repeated/unique), presentation (enhanced, normal), foil type (repeated/unique), and their interactions as well as key allocation as fixed effects, and the by-subjects and by-items intercepts and slopes for target type, presentation and foil type as random effects. The interactions between the factors were not included as random factors as the full model failed to converge. The effects of target type and foil type were significant, $\chi^2(1) = 63.434$, $p < .001$, $\chi^2(1) = 328.123$, $p < .001$ respectively, as was their interaction $\chi^2(1) = 14.815$, $p < .001$. Pairwise comparisons between the repeated and unique target type conditions revealed a significant difference in both wrong repetition and substitution foil type conditions, $\chi^2(1) = 74.436$, $p < .001$; $\chi^2(1) = 13.092$, $p < .001$. Participants were significantly less accurate when they had to reject a foil for the repeated target type than for the unique target type. The interaction between foil type and presentation was also significant, $\chi^2(1) = 3.849$, $p = .050$, as was the interaction between foil type and key allocation, $\chi^2(1) = 3.858$, $p = .050$. The three-way interaction between key allocation, presentation type and foil type was not significant, $\chi^2(1) = 2.334$, $p = .127$. There were no other significant results.

Table 12. Mean Accuracy (%) by Condition in Different Trials, Experiment 4. Examples of repeated and unique letter targets with corresponding substitution and wrong repetition foils: OLELUVAN - OL**YM**UVAN/O**E**LEUVAN; OBELUVAN - OB**YM**UVAN/O**E**BEUVAN

	Foil type			
	Unique (substitution)		Repeated (wrong repetition)	
	Presentation			
Target type	Normal	Enhanced	Normal	Enhanced
Repeated	60.5	64.3	31.0	29.8
Unique	66.5	66.3	42.9	41.7

6.4. Modelling

6.4.3. LTRS

Adapting LTRS to same-different decisions

In Experiment 4, a brief stimulus *s* is presented, before a mask and a clearly presented probe *w* for a same-different judgement between the two stimuli. This does not correspond to any previous application of LTRS, but there is a straightforward extension, which we will describe.

On same trials $v(w)$ will always be 1 whereas on different trials $v(w)$ could be 0 or 1. Thus a v of 0 is unambiguous evidence for "different" so responding is accurate when v is 0 unless a premature guess is made with probability ε. In contrast, a v of 1 is ambiguous, unless all x_i are 2, in which case the percept is complete and consistent with the comparator; this means a "same" response must be correct, and this response will be made unless a premature guess has occurred. Otherwise, a guess is made, which is a "same" response with probability g_s. Thus, accuracy on different trials is computed:

P(no guess | different trial) = $(1-\varepsilon)P(V(w)=0; s, t, \text{masked})$

P(accurate | different trial) =

P(no guess | different trial) + P(the guess is different)P(guess | different trial) =

$$(1-\varepsilon)P(V(w)=0; s, t, \text{masked}) + (1-g_s)(1-(1-\varepsilon)P(V(w)=0; s, t, \text{masked})) \quad (10)$$

and accuracy on same trials is computed thus:

P(no guess | same) = $(1-\varepsilon)\prod_i P(X_i=2; t)$

P(accurate | same trial) =

P(no guess | same trial) + P(the guess is same)P(guess | same trial) =

$$(1-\varepsilon)\prod_i P(X_i=2; t) + g_s\left(1-(1-\varepsilon)\prod_i P(X_i=2; t)\right) \quad (11)$$

Provable properties of LTRS predictions

A detailed enumeration of the ***x*** states for the different trials is given in Appendix D along with an explanation of how these can (or cannot) derive predictions for the simple effects and interaction for these trials (ignoring the enhancement manipulation). These show that the predictions of LTRS are that: There is a simple effect of target repetition to create more guesses and hence more errors for both foil conditions (because these probabilities must be positive) and the interaction is such that the target repetition effect is greater when there is also a repetition in the foil. Some details follow.

Repetition in the target The source of the effect of repetitions in the target is that the unique target receives accurate responses in cases of ***x*** that perceives an identity that does not appear in the foils (i.e., a unique identity) and where in the corresponding case for the repeated target, a single twinned letter is perceived without other inconsistency information being unavailable causing a guess. In other words, either of two letters in the repeated target potentially match one letter in the foil whereas the same letter in the foil is matched by only one letter of the unique target.

Repetition in the foil We cannot state a parameter-free prediction of the model for the effect of foil repetition because while there are several cases of $\boldsymbol{x}$ where the construction of the repetition foil reduces accuracy, there are also a few cases where the reordering of letters in the repeated-letter foil increases accuracy.

When the letter between the critical letter and its (potential) twin is perceived, this produces an inconsistency in the unique foil but not necessarily the repetition foil, as this letter is present (indeed, present twice) in the repetition foil but not (at all) in the unique foil. There are many such cases that give an accuracy advantage to the unique foil conditions.

There are, however, a limited number of cases that provide a countermanding influence (give an accuracy disadvantage to the unique foil conditions). These involve the critical letter and its outside flanker being perceived fully and the potential twin and the interior letter not being perceived at all. It is presently unclear whether it is possible to find parameters where these cases outweigh the cases with the opposite effect, but numerical fits are presented below. In any case, according to LTRS, the effect of foil repetition here has nothing to do with repetition per se, and everything to do with letter identities and letter orderings being preserved.

Interaction There is a substring of the target, involving the middle letter and the twin or potential twin of the critical letter, that is only present in the foil when both target and foil have repeated letters (e.g., target: OLELUVAN, foil: OELEUVAN have the shared bigram “EL” but OLELUVAN-OLYMUVAN, OBELUVAN-OEBEUVAN and OBELUVAN-OBYMUVAN do not contain any such shared bigrams in the letter triple that is manipulated). This provides an additional opportunity for confusion and reduces accuracy predicted by LTRS for this cell of the design only. This creates an interaction involving the two repetition effects compatible with the data, but again, LTRS is not attributing this effect to repetition per se.

Numerical Predictions of LTRS

Mindful of the risks of overfitting, we again fixed some parameters and constrained others to be equal. The parameters that were fixed mostly control the effect of stimulus duration, which was not manipulated in this experiment, and those that were constrained were for different positions, which we are averaging over anyway, so these restrictions are not important for representing how the model performs for these data. For the simulation of Experiment 4, the implementation for perceptual identification task in LTRS was changed so that it reflects the same-different nature of the task in this experiment, and the enhancement manipulation was ignored. Unlike the original implementation of the identification task with two-alternative forced choice, where the decision is determined mostly by comparing the compatibility of the two alternatives with the percept (and the correct option is always compatible), here the decision is made by comparison between the percept and the comparator. Two changes were made. First, on same trials, the perfect match between target and comparator is positively determined if (and only if) all perceptual elements are available. Second, for trials where neither a positive determination of 'same' or 'different' is available, then the bias parameter is treated as the probability of a 'same' response. This implementation, therefore, is not able to capture any effects in the same trials as it will produce identical results in all same trial conditions when the perceptual parameters are fixed. We examined predictions for different trials[6] after optimizing letter rates (subject to constraints)[7] and 'same'-response bias parameters on the basis of all trials. As can be observed in Figure 14, LTRS managed to capture the data pattern, specifically, the lower accuracy in cases of repeated targets especially when the target was compared with the repeated letter foil type.

6.4.4. Overlap Model

To adapt the Overlap Model for the task with a single comparator string, the overlap for the "missing" alternative was replaced with a parameter *t* that operated as a form of soft threshold. That is:

$$P(same) = \frac{O(T_k, S)^{a_k}}{O(T_k, S)^{a_k} + t} \tag{12}$$

If the overlap between target and reference, raised to the relevant exponent, is equal to the threshold, then the model is guessing between "same" and "different" responses, with greater overlap increasing the probability of a "same" response.

[6] A stronger 'same'-response bias for trials with repeated letters in the comparator would suffice to accommodate the 'same' trials without disrupting the model's ability to make the other predictions correctly (it would enhance the foil-type difference) but is not interesting.
[7] Here, interior letter rate parameters were fixed to be equal, and parameters not optimized were chosen arbitrarily. The full parameter set is given in the appendix.

A simulation was run with the Overlap Model after optimization of the following parameters: scaling exponent for a unique condition (a), scaling exponent for a repeated condition (b), the asymptote and rate parameters (controlling the standard deviations for the letter positions), and a threshold value. The optimized parameters are listed in appendix B. As can be seen in Figure 14*,* the Overlap Model managed to match the observed data pattern quite well and predicted the presence of repetition effects, with bigger magnitude in the condition with wrong repetition foils.

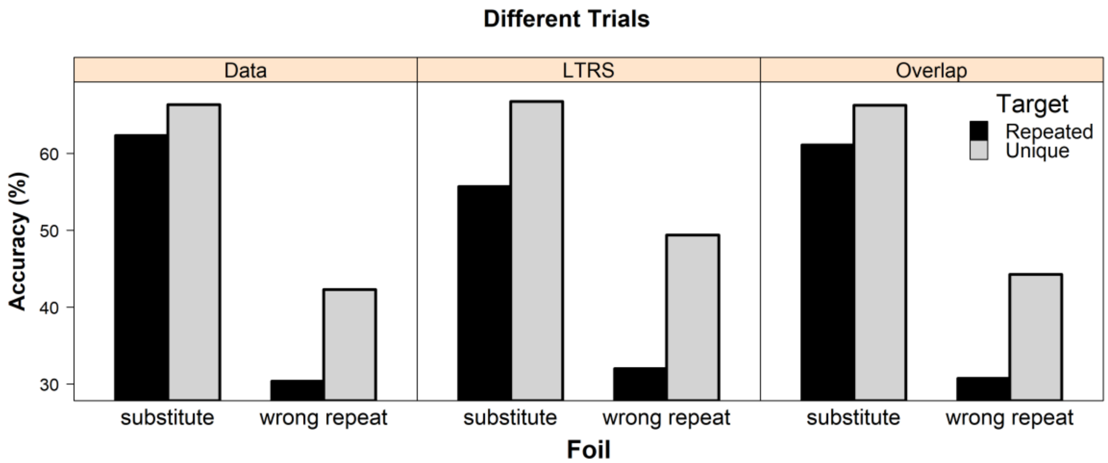

Figure 14. Mean accuracy by target type and foil type in different trials, Experiment 4. Empirical data (left), LTRS accuracy predictions (middle), and Overlap accuracy predictions (right). Examples of repeated and unique letter targets with corresponding substitution and wrong repetition foils: OLELUVAN-OL**YM**UVAN/O**E**LEUVAN; OBELUVAN - OB**YM**UVAN/O**E**BEUVAN

5.3.3. Open bigram models and SOLAR

As in the insertion and deletion conditions in Experiments 1-3, we explored the predictions of the SOLAR model (Davis, 1999), as well as the two versions of the overlap bigram models (Grainger et al., 2006; Grainger & van Heuven, 2003) by calculating match scores for the substitution and wrong repetition foils in Experiment 4, separately for each target type. As can be seen in Table 13, SOLAR again predicted no difference between the repeated and the unique conditions for each of the foil types. In addition, the difference between the similarity of both foil types with the target was predicted to be small. The open bigram models, on the contrary, predicted stronger similarities between the repeated letter targets and the two foil types than between the unique letter target and the same foil types. In addition, this difference was particularly large in the cases of the wrong repetition foils, which were also predicted to be generally much more similar to both target types than the corresponding substitution foils. The open bigram schemes were in accordance with the observed data pattern.

Table 13. Mean match score calculations by target type and foil type in Experiment 4 for SOLAR (Davis, 1999), Open Bigram (Grainger & van Heuven, 2003), Overlap Open Bigram (Grainger et al., 2006)

Foil Type	**Substitution** (OLYMUVAN / OBYMUVAN)		**Wrong Repetition** (OELEUVAN / OEBEUVAN)	
Target Type	**Repeated** (OLELUVAN)	**Unique** (OBELUVAN)	**Repeated** (OLELUVAN)	**Unique** (OBELUVAN)
Model				
SOLAR	0.8	0.8	0.85	0.85
Open Bigram	0.52	0.46	0.88	0.68
Overlap Open Bigram	0.54	0.53	0.79	0.73

6.5. Discussion

The results of Experiment 4 provided strong evidence that the presence of a repetition changes the mechanisms underlying the processing of a letter string. It also demonstrates that the results are task dependent. In the same trials, participants were facilitated when the target contained a repetition and the number of possible identities was one less than the condition in which the target contained eight unique identities. The results in the different trials were in the opposite direction and the repeated letter conditions were significantly harder than the unique conditions. The fact that both foil types were harder to reject when compared to repeated letter targets than unique letter ones suggests that the presence of repeated letters makes the string difficult to distinguish and generally more similar to many input candidates. This means that the result of Experiment 1 that there was no cost for repeated letters in briefly presented targets does not generalize. The pattern of the results in the different trials was captured in the predictions of LTRS (Adelman, 2011) and the Overlap Model (Gomez et al., 2008) and were also in accordance with repetition-sensitive open bigram schemes (Grainger et al., 2006; Grainger & van Heuven, 2003). The results disagreed with the predictions of SOLAR (Davis, 1999).

7. General Discussion

The results from these experiments provide strong evidence that the mechanisms underlying processing of repeated letters in a letter string are different than those of unique letters. Repetition effects were observed in three paradigms contrasting nonadjacent repetition of letters with unique letters: a two-alternative forced-choice perceptual identification task, a masked primed same-different task, and a same-different perceptual identification task. The results also suggested that the pattern of the effect was sensitive to methodological idiosyncrasies and should be interpreted in the context of the experimental task. Table 14 summarizes the key experimental results and the predictions of models discussed below.

Table 14. Summary of the data pattern in Experiments 1-4 and corresponding predictions of visual word recognition models

Experiment	Data Pattern	LTRS	Overlap	Open Bigram	SOLAR
E1: 2AFC Perceptual Identification Lower accuracy (<) suggests higher similarity with the target	Insert in Foil: Repeat < Unique Delete in Foil: Repeat = Unique	As data (model with repetition bias)	As data	Predicts difference in accuracy for both insert and delete: Repeat < Unique	Predicts no difference in both insert and delete: Repeat = Unique
E2: Masked-priming Same-Different (same trials) Larger priming (>) suggests higher similarity with the target	Delete in Prime: Repeat = Unique	As data	Predicts difference in priming: Repeat > Unique*	Predicts difference in priming: Repeat > Unique	As data
E3: Masked-priming Same-Different (same trials) Larger priming (>) suggests higher similarity with the target	Insert in Prime: Repeat > Unique	As data	As data	As data	Predicts no difference: Repeat = Unique
E4: Same-Different Perceptual Identification (different trials) Lower accuracy (<) suggests higher similarity with the target	Substitute in Foil: Repeat < Unique Wrong Repeat in Foil: Repeat < Unique	As data	As data	As data	Predicts no difference for both: Repeat = Unique

Note – *Although match scores were higher for primes containing repetition than primes containing unique letters, the overlap of the targets with themselves was also higher for targets with repeated letters than targets with only unique letters, leading to difficulty in interpreting the predicted results.

Experiment 1 used a two-alternative forced-choice perceptual identification paradigm. The foils had an additional or missing letter relative to the target, and the inserted or deleted letter could be a repetition of another target letter identity or a unique identity. A letter repetition effect was observed such that foils with an inserted repetition were more likely to be incorrectly chosen than foils with an inserted unique letter. This suggests a difficulty with identifying the numerosity of letters in the briefly presented target string: Participants chose foils with the wrong number of a particular letter more than they chose foils with a wrong letter. There was no similar robust effect for comparisons of targets with and without repetitions. In contrast, the basic mechanisms of LTRS (Adelman, 2011) predict that ambiguity is created when the briefly presented targets have repeated letters relative to unique letters: A letter identity present in the target but not the foil is more easily perceived that the numerosity of the repetition or other disambiguating positional information. At a verbal level of description, this agrees with our characterization of the data: Numerosity detection is difficult, so decisions where one of the options contains a repetition are difficult.

With the original decision rule of LTRS, though, the model produces a repetition effect in the foil-deletion trials, not the foil-insertion trials, because those trials whose foils omit a repeated letter have targets that contain the difficult-to-perceive repetition. However, participants could try to counteract this difficulty by using a different decision rule that disproportionately chooses the repeated letter option when such an option is available and the percept is ambiguous. Additional modelling with LTRS with this extra response bias mechanism produced the observed pattern of results.

Gomez et al. (2008) had also found it necessary to include a response bias involving repeated letters in their modelling of two-alternative forced-choice data with the Overlap Model, but they required the opposite bias to avoid choosing the option with the repeated letter. Such a bias would increase accuracy for foils that include repetitions, as these would be rejected more often, but in the data, these were numerically the least accurate trials. Nevertheless, the Overlap Model was successful in modelling the outcome of the present Experiment 1, regardless of whether the extra parameter for bias was used or not.

The results of Experiment 1 were more consistent with the predictions of open bigram models (Grainger et al., 2006; Grainger & van Heuven, 2003) which predict higher similarity in the conditions with repeated letters than conditions with unique letter only, than SOLAR, which predicts that no repetition effect should be observed as the two letter conditions are equally similar (Davis, 1999). This experiment provided evidence that repeated letter identities are not always processed as unique identities, therefore challenging the encoding scheme of SOLAR.

The same insertion and deletion manipulations were used to explore repeated letter effects with a masked priming methodology in the next two experiments. The more implicit measurement of confusability in these paradigms – participants have limited awareness of, and no explicit task on, the prime – means that the results can be considered more directly, without consideration of response biases.

If items that differ only in numerosity of a letter are more confusable than those differing in a letter identity, then it would be expected that primes with a letter deleted relative to the target would be more effective if that deletion changed only the numerosity of a letter – that is, deleted a repeated letter. No such effect was found (on "same" responses) in our Experiment 2 with a masked primed same-different task; this was consistent with the findings of Schoonbaert and Grainger (2004) with a similar manipulation in the masked primed lexical decision task.

If detecting the numerosity of a letter in a briefly presented string posed a particular difficulty, then it would be expected that primes with a letter inserted relative to the target would be more effective if that insertion changed only the numerosity of a letter – that is, the insertion created a repeated letter. In Experiment 3, this effect was found. Given the implicit nature of the priming manipulation, we expect that this insertion-deletion asymmetry in repetition effects – observed without relevant response biases – is the one that naturally arises from orthographic processing. We believe the finding of Experiment 3 to be novel: Van Assche and Grainger's (2006) previous investigation of primes with insertions did not find a difference between repeated and unique insertions. However, Van Assche and Grainger used masked primed lexical decision, which Norris et al. (2010) have argued is less sensitive than the masked primed same-different task that we used.

If the presence of a repetition effect in Experiment 3 but not Experiment 2 could imply the effect will only emerge when the manipulation of repetition affects the briefly presented string, but the opposite pattern was seen in Expeirment 1. There, the effect emerged only when the manipulation affected the foil, and the options presented effectively required a decision on letter numerosity. The apparent inconsistency might therefore be attributed to properties of the task. This was reflected in our modelling with LTRS. While for Experiment 1, LTRS needed a task-specific process to accommodate the findings, LTRS naturally makes the correct prediction for Experiments 2 and 3. A repetition effect occurs in LTRS only for the insertion primes and not the deletion primes, because the insertion creates an inconsistent piece of information that will require more processing to detect in the repeated case, but the deletion leaves behind the same inconsistency regardless of repetition. The same reasoning – that it is the presence of inconsistent information, not the absence of consistent information, that controls priming – applies to comparisons of replacement (zudge-JUDGE) and insertion (zjudge-JUDGE) primes: LTRS performance is largely controlled by inconsistent (z) information not consistent (j) information, and so these types of primes should be equivalent, as Lupker, Spinelli and Davis (2020) have found.

The representational scheme in the Overlap Model allows for there to be more similarity in representations involving letter repetitions than those without, because representations of letters leak into nearby positions, allowing the letters in repetitions to contribute to the overlap more than once. This would account for the findings of a repetition effect in Experiment 3, but not the absence of one in Experiment 2, though there is some ambiguity here, because for the between-target comparison, the targets had different maximum similarities, and the model does not offer an explicit account of this kind of priming task.

As before, representations based on open bigrams gave higher similarities for the stimulus pairs involving repetitions, suggesting a repetition effect should be observed, and SOLAR representations treated repetitions as irrelevant, suggesting none should be observed. While neither of these types of models predicted the strict asymmetry in repetition effects, all suggest that there is an asymmetry between insertions and deletions, which may contribute to the different effect of repetitions on insertions compared to deletions.

While Schoonbaert and Grainger's (2004) experiments and Trifonova and Adelman's (2019) megastudy analyses found that lexical decisions to words were slower for items with letter repetitions, in Experiment 2's same-different task with nonwords, latencies to respond "same" were not affected by letter repetition. The difference between the prior studies and the present Experiment 2 may be due to task-specific processes linked to lexical decision and the same-different task, or to differences in the type of stimuli. For instance, if the repetition effect on clearly presented stimuli is based on how well those stimuli have been learned – which could be affected by perceptual difficulty due to repetition – then the effect would be limited to words.

Experiment 4, which examined the accuracy of identification of briefly presented strings with a same-different decision, also did not find an inhibitory effect of repeated letters on "same" trials; indeed, the effect was that "same" received more accurate responses when a repetition was present. This effect was anticipated by the Overlap Model, which allows for repeated letters to contribute more than once to overlap calculations. By contrast, LTRS would again need to include a response bias for the repeated letter items to account for this finding.

In the "different" trials of Experiment 4, though, less accurate responding was seen for targets involving repetitions. This effect was stronger when the foil contained a repetition that was not in the target (and removed the original repetition if present, so unique: OBELUVAN-OEBEUVAN; repeated: OLELUVAN-OELEUVAN) than it was when the foil replaced two letters (including a repeated one if present, so unique: OBELUVAN-OBYMUVAN; repeated: OLELUVAN-OLYMUVAN). The pattern of performance on different trials was anticipated by LTRS and Overlap Model, and by the open bigram representational schemes. Nevertheless, this is not a universal prediction. In edit distance schemes of similarity, both foils with repetition are distant by one deletion and one insertion; and foils with two substitutions are distant by those two substitutions. Moreover, as before, the SOLAR scheme is generally insensitive to repetition.

Across our experiments, it appears that unique identities in letter strings are more easily processed accurately than repetitions, and strings with repetitions are therefore treated as more similar to comparison strings. While the duplication may offer an advantage in detecting the presence of the repeated letter identity, identifying the numerosity of repeated identity seems to introduce processing obstacles that do not occur when the same number of distinct identities need to be processed. Future research is needed to establish whether the observed effects are specific to letter strings or also generalize to other domains, such as digit strings, as well.

The evidence provided in the present study suggests that there are processing limitations in the encoding of repeated letter information in strings. These results are in accordance with findings of tachistoscopic letter identification studies (Bjork & Murray, 1977; Egeth & Santee, 1981) demonstrating interference in cases of processing two identical letters in a visual array, the study of Mozer (1989) demonstrating lower accuracy in evaluation of numerosity of letters in cases of repetitions, and resemble the perceptual phenomenon of repetition blindness (Kanwisher 1987; 1991; Luo & Caramazza, 1996).

Taken together, the evidence provided by the experiments here was broadly consistent with LTRS (Adelman, 2011) and its predictions that letter repetition in strings requires additional processing steps for disambiguation. It was also generally in line with models that could account for repetition effects due to mechanisms in which letter position is weakly represented, such as the Overlap Model (Gomez et. al, 2008) and open bigram models (Grainger et al, 2006; Grainger & van Heuven, 2003). The data, however, also presented some challenge to all aforementioned modelling frameworks. The observed repetition effects are at odds with schemes, such as the spatial coding, that treat unique letters in the same way as repeated ones (SCM, SOLAR; Davis, 1999; 2010). The results are also not entirely consistent with open bigram schemes which would suggest that there should be difference between the repeated and unique conditions in both insertion and deletion cases. Although the predictions of both LTRS and the Overlap Model generally captured the data pattern well, LTRS required an additional implementation of repetition bias to fit the data of Experiment 1 and the match scores for Experiment 2 of the Overlap Model could be interpreted as prediction for repetition effect in both identity and deletion primes where there was none.

This study extends previous findings of perceptual effects of repetition and demonstrates repeated letter effects when letters are embedded in strings. The results would be interpreted in many frameworks as suggesting that string stimuli containing repeated letters bear close resemblance to more input candidates than corresponding unique letters strings. The special properties of letter repetitions highlight to us, however, that assumptions regarding the relationship between letter identity and letter position – and indeed the concept of resemblance – require critical consideration. Since letter identities do not always have a one-to-one correspondence to letter positions, it is not necessarily natural to consider the letter location as an integral part of the same component representation as the letter identity.

Most models of orthographic processing treat information about letter identity and information about letter position as part of an integrated orthographic code that becomes stronger and/or clearer over time. The influence of both identity and position on lexical processing increases over time due to a single integration (e.g., activation) process that continually compares a stored orthographic code to a perceived orthographic code. In these models, the influence of identity may be stronger than that of position. This means that when time is short, we may not have the power to experimentally detect a small influence of position but may have the power to detect stronger influence of identity. However, the sense here in which positional information can be "slower" than identity is only illusory: The positional effect is theoretically present but undetectable.

However, it seems perfectly feasible and natural to ask whether the information about positions of letters – or even numerosities of letters – becomes available to lexical identification processes later than information about the presence of letter identities. In LTRS (Adelman, 2011), positional information is available later in part directly and in part indirectly because the order of two letters requires both to have been perceived. It is by no means the only conceivable model in which different perceptual dimensions of letter strings become available at different times. We believe exploring the variety of possible models of this type and the corresponding empirical questions are likely to be more productive than positing increasingly complex match score calculations as underpinning orthographic processing.

Allowing for different perceptual information to become available for consideration at different times is critically a possible mechanism to dissociate salience from importance. Salient things are generally perceived sooner, but they are not necessarily the most relevant to a perceptual decision. Therefore, it seems possible that information perceived late could be given any amount of weight (i.e, more or less than information perceived early) in deciding the identity of a word. This is not possible if resemblance drives both apparent perceptual speed and lexical matching. As a matter of rhetoric (and mathematical simplification) LTRS (Adelman, 2011) totally dissociates salience and importance by keeping importance constant – all negative information is critically important – but this is not a necessary part of this dissociation: LTRS's relatives in categorization (Lamberts, 2000, 2002) parameterize decisional weight separately from salience.

Placing the present results alongside related ones concerning delayed processing for words with repeated letters (Schoonbaert & Grainger, 2004; Trifonova & Adelman, 2019) suggests that both are linked to issues of salience of positional information for repeated letters in perception. We anticipate that it will prove important to conceptually distinguish the concept of salience as perceptual speed from the concepts of decisional weight and attention. Progress in this direction may depend on models becoming more general – in terms of tasks – as new paradigms attempt to probe these issues more directly.

Acknowledgments

This work was supported by the Leverhulme Trust (Project Grant RPG-2013-408) and an Early Career Fellowship awarded to the first author by the Institute of Advanced Study, University of Warwick, United Kingdom.

References

Adelman, J. S. (2011). Letters and Time in Retinotopic Space. *Psychological Review, 118,* 570–582.

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. https://doi.org/10.3758/s13428-013-0442-y

Aschenbrenner, A. J., Balota, D. A., Weigand, A. J., Scaltritti, M., & Besner, D. (2017). The first letter position effect in visual word recognition: The role of spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 700–718. https://doi.org/10.1037/xhp0000342

Bjork, E. L., & Murray, J. T. (1977). On the nature of input channels in visual processing. *Psychological Review*, *84*, 472–484.

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.

Davis, C. J. (1999). The self-organizing lexical acquisition and recognition (SOLAR) model of visual word recognition. Unpublished doctoral dissertation, University of New South Wales, Sydney, New South Wales, Australia.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological review*, *117*, 713.

Egeth, H. E., & Santee, J. L. (1981). Conceptual and perceptual components of interletter inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 506.

Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology Section A, 39*, 211–251. https://doi.org/10.1080/14640748708401785

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods*, *35*, 116-124.

Gomez, P. (2020, July 15). *Overlap Model Code*. https://doi.org/10.31234/osf.io/67g2f

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

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

Grainger, J., & van Heuven, W. J. B. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The mental lexicon* (pp. 1–23). New York: Nova Science.

Jordan, T. R., Thomas, S. M., Patching, G. R., & Scott-Brown, K. C. (2003). Assessing the importance of letter pairs in initial, exterior, and interior positions in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 883.

Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition, (27)*, 117-143.

Kanwisher, N. G. (1991). Repetition blindness and illusory conjunctions: Errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 404.

Kinoshita, S., & Kaplan, L. (2008). Priming of abstract letter identities in the letter match task. *Quarterly Journal of Experimental Psychology*, *61*, 1873-1885.

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

Lamberts, K. (2000). Information accumulation theory of speeded categorization. *Psychological Review, 107*, 227-260.

Lamberts, K. (2002). Feature sampling in categorization and recognition of objects. *Quarterly Journal of Experimental Psychology, 55A*(1), 141-154.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis.* New York: Wiley

Luo, C. R., & Caramazza, A. (1996). Temporal and spatial repetition blindness: Effects of presentation mode and repetition lag on the perception of repeated items. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 95.

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*, 1281–1299. https://doi.org/10.1037/xhp0000087

Lupker, S. J., Nakayama, M., & Yoshihara, M. (2018). Phonologically-Based Priming in the Same-Different Task With L1 Readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 44*, 1317.

Lupker, S. J., Perea, M., & Davis, C. J. (2008). Transposed-letter effects: Consonants, vowels and letter frequency. *Language and Cognitive Processes, 23*, 93–116. https://doi.org/10.1080/01690960701579714

Lupker, S. J., Spinelli, G., & Davis, C. J. (2020). Is zjudge a better prime for JUDGE than zudge is?: A new evaluation of current orthographic coding models. *Journal of Experimental Psychology: Human Perception and Performance.* Advance online publication. https://doi.org/10.1037/xhp0000856

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: An account of basic findings. *Psychological Review, 88,* 375–407. http://dx.doi.org/10.1037/0033-295X.88.5.375

Mozer, M. C. (1989). Types and tokens in visual letter perception. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 287–303. https://doi.org/10.1037//0096-1523.15.2.287

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

Norris, D., Kinoshita, S., & van Casteren, M. (2010). A stimulus sampling theory of letter identity and order. *Journal of Memory and Language*, *62*, 254–271. https://doi.org/10.1016/j.jml.2009.11.002

Perea, M., & Lupker, S. J. (2003). Transposed-letter confusability effects in masked form priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 97–120). Hove, UK: Psychology Press.

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

Peressotti, F., & Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception & Psychophysics, 61*, 691–706.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, *88*, 552.

Sayim, B., & Taylor, H. (2019). Letters Lost: Capturing Appearance in Crowded Peripheral Vision Reveals a New Kind of Masking. *Psychological Science, 30*, 1082–1086. https://doi.org/10.1177/0956797619847166

Scaltritti, M., & Balota, D. A. (2013). Are all letters really processed equally and in parallel? Further evidence of a robust first letter advantage. *Acta Psychologica*, *144*, 397–410. https://doi.org/10.1016/j.actpsy.2013.07.018

Scaltritti, M., Dufau, S., & Grainger, J. (2018). Stimulus orientation and the first-letter advantage. *Acta Psychologica*, *183*, 37–42. https://doi.org/10.1016/j.actpsy.2017.12.009

Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes, 19*, 333–367. https://doi.org/10.1080/01690960344000198

Trifonova, I. V., & Adelman, J. S. (2019). A delay in processing for repeated letters: Evidence from megastudies. *Cognition*, *189*, 227-241.

Tydgat, I., & Grainger, J. (2009). Serial position effects in the identification of letters, digits, and symbols. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 480–498. https://doi.org/10.1037/a0013027

Van Assche, E., & Grainger, J. (2006). A study of relative-position priming with superset primes. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 32*, 399–415. https://doi.org/10.1037/0278-7393.32.2.399

Welvaert, M., Farioli, F., & Grainger, J. (2008). Graded Effects of Number of Inserted Letters in Superset Priming. *Experimental Psychology, 55*, 54–63. https://doi.org/10.1027/1618-3169.55.1.54

Appendix A. LTRS (Adelman, 2011) Parameters

Table A1. LTRS parameters for Experiment 1, original implementation

Parameter	a	sd	initial8	other8	initial7	other7	ee	phi	lambda
Value	20	10	0.118	0.004	0.026	0.025	0.01	0.01	1

Table A2. LTRS parameters for Experiment 1, implementation with repeated letter bias

Parameter	a	sd	initial8	other8	initial7	other7	ee	phi	lambda	rl
Value	20	10	0.278	0.003	0.027	0.027	0.01	0.01	1	0.082

Table A3. LTRS parameters for Experiment 4

Parameter	a	sd	initial	middle	final	ee	phi	lambda	ps
Value	38	20	0.87237	0.00553	0.00001	0.01	0.01	0.03	0.783

Appendix B. Overlap Model (Gomez et al., 2008) Parameters

Table B1. Overlap parameters for Experiment 1

Parameter	Overlap with bias	Overlap without bias
a	6.776	6.708
b	6.803	-
d	1.007	1.007
r	1.011	1.072

Table B2. Overlap parameters for Experiment 4

Parameter	a	b	d	r	threshold
Value	7.784	7.717	1.284	1.599	3442.776

Appendix C.

Tabulation of states to analyse conditions of Experiments 1-3

Table C1. Map from perceptual states $\boldsymbol{x}$ to consistency status v for briefly presented 7-letter stimuli (insertion-in-foil conditions for Experiment 1; deletion-prime conditions in Experiment 2).

$\boldsymbol{x}$ constraint	v unique-f	v repeat-f	Unique example s: dratien f or w: DRALTIEN	Repeat example s: dratien f or w: DRARTIEN
	$v(f)$ (foil: Exp. 1) or $v(w)$ (target: Exp. 2)		i = 3 (letter a); i + 1 = 4 (letter t)	
$x_i = 2$; $x_{i+1} = 2$	0	0	at	at
$x_i = 2$; $x_{i+1} = 1$	1	1	a*t	a*t
$x_i = 2$; $x_{i+1} = 0$	1	1	a*/a+	a*/a+
$x_i = 1$; $x_{i+1} = 2$	1	1	a*t	a*t
$x_i = 1$; $x_{i+1} = 1$	1	1	a*t	a*t
$x_i = 1$; $x_{i+1} = 0$	1	1	a*	a*
$x_i = 0$; $x_{i+1} = 2$	1	1	*t/+t	*t/+t
$x_i = 0$; $x_{i+1} = 1$	1	1	*t	*t
$x_i = 0$; $x_{i+1} = 0$	1	1	*/+	*/+t

Note – *i* refers to the position before the missing letter and *i* + 1 refers to the position after the deletion. There is no representation of the irrelevant initial dr or final ien letters; the r is irrelevant because there is nothing about it that can affect the match: its presence, postposition to d, preposition to a, and relative location to the non-adjacent letters are all consistent with both the *w* options. A slash (/) means "or".

Table C2. Map from perceptual states $\boldsymbol{x}$ to consistency status v for briefly presented 8-letter stimuli (deletion-in-foil conditions for Experiment 1; insertion-prime conditions in Experiment 3)

$\boldsymbol{x}$ constraint	v unique-s	v repeat-s	Unique example s: draltien f or w: DRATIEN	Repeat example s: drartien f or w: DRATIEN
	$v(f)$ (foil: Exp. 1) or $v(w)$ (target: Exp. 3)		i = 4 (l or r); j = 3 (a); k = 2 (r); h = 5 (t)	
$x_i > 0; x_k > 0$	0	0	Contains l (and r)	Contains r twice
$x_h = 2; x_i = 0; x_j = 2$	0	0	*a+t/r*a+t	*a+t/r*a+t
$x_h = 2; x_i > 0; x_j > 0; x_k = 0$	0	0	*alt/*a*lt/*a*l*t	*art/*a*rt*/*a*r*t*
$x_h = 1; x_i > 0; x_j > 0; x_k = 0$	0	0	*a*l*t/*al*t	*a*r*t*/*ar*t*
$x_h = 2; x_i = 2; x_j = 0; x_k = 0$	0	0	*lt	*rt
$x_h = 2; x_i = 1; x_j = 0; x_k = 0$	0	1	*l*t	*r*t
$x_h < 2; x_i > 0; x_j = 0; x_k = 0$	0	1	*l*/*l*t	*r*/*r*t
$x_h = 2; x_i = 0; x_j = 1$	1	1	*a*t/r*a*t	*a*t/r*a*t
$x_h = 1; x_i = 0; x_j > 0; x_k > 0$	1	1	ra*t/r*a*t	ra*t/r*a*t
$x_h = 1; x_i = 0; x_j > 0; x_k = 0$	1	1	*a*t/+a*t	*a*t/+a*t
$x_h = 1; x_i = 0; x_j = 0; x_k > 0$	1	1	r*t	r*t
$x_h = 0; x_i = 0; x_j > 0; x_k > 0$	1	1	ra*/r*a*	ra*/r*a*
$x_h = 0; x_i = 0; x_j > 0; x_k = 0$	1	1	*a*/+a*	*a*/+a*
$x_h = 0; x_i = 0; x_j = 0; x_k > 0$	1	1	r*/r+	r/r+
$x_h > 0; x_i = 0; x_j = 0; x_k = 0$	1	1	*t	*t
$x_h = 0; x_i = 0; x_j = 0; x_k = 0$	1	1	*/+	*/+

Note – i refers to the position in f or s of the deleted letter; k refers to the position of the repetition (or where the repetition would be); and j refers to the intermediate position; and h refers to the position adjacent to i that is not j. There is no representation of the irrelevant d or ien letters in the examples. A slash (/) means "or".

All the cases where the two v columns differ, the unique column has the advantage (is the one in which the inconsistency is detectable), and so the unique foil has greater probability of inconsistency, regardless of the exact probabilities of those rows. Therefore, the model categorically predicts (regardless of parameters or stimulus duration) that responses will be more accurate for these foils (if guessing is equally accurate).

Appendix D.

Tabulation of states to analyse conditions of Experiment 4

Table D1. Map from perceptual states $\boldsymbol{x}$ to consistency status v for briefly presented stimuli and their foils on "different" trials in Experiment 4.

$\boldsymbol{x}$ constraint	Repetition in target (e.g., OLELUVAN)		No repetition in target (e.g, OBELUVAN)		fx
	Rep. in foil OELEUVAN	No foil rep. OLYMUVAN	Rep. in foil OEBEUVAN	No foil rep. OBYMUVAN	
$x_i > 0; x_k > 0$	0	0	0	0	*N*
$x_k = 2; x_l = 2$	0	0	0	0	*N*
$x_h = 2; x_i = 2; x_j > 0; x_k = 0$	0	0	0	0	*N*
$x_h < 2; x_i = 2; x_j = 1; x_k = 0$	1	0	1	0	*F*
$x_h = 2; x_i = 2; x_j = 0; x_k = 0$	0	1	0	1	*f*
$x_h < 2; x_i = 2; x_j = 0; x_k = 0; x_l < 2$	1	1	1	1	*A*
$x_i = 1; x_j = 2; x_k = 0; x_l = 2$	0	0	0	0	*N*
$x_i = 1; x_j = 2; x_k = 0; x_l < 2$	1	0	1	0	*F*
$x_i = 1; x_j = 1; x_k = 0$	1	0	1	0	*F*
$x_i = 1; x_j = 0; x_k = 0$	1	1	1	1	*A*
$x_h = 2; x_i = 0; x_j = 2; x_k = 2; x_l < 2$	0	0	0	0	*N*
$x_h < 2; x_i = 0; x_j = 2; x_k = 2; x_l < 2$	1	0	0	0	*I*
$x_h = 2; x_i = 0; x_j = 2; x_k = 1$	0	0	0	0	*N*
$x_h < 2; x_i = 0; x_j = 2; x_k = 1$	1	0	0	0	*I*
$x_h = 2; x_i = 0; x_j = 2; x_k = 0; x_l = 2$	0	0	0	0	*N*
$x_h = 2; x_i = 0; x_j = 2; x_k = 0; x_l < 2$	1	0	1	0	*F*
$x_i = 0; x_j = 1; x_k = 2; x_l < 2$	1	0	0	0	*I*
$x_i = 0; x_j = 1; x_k = 1$	1	0	0	0	*I*
$x_h = 2; x_i = 0; x_j = 1; x_k = 0$	1	0	1	0	*F*
$x_h < 2; x_i = 0; x_j > 0; x_k = 0$	1	0	1	0	*F*
$x_i = 0; x_j = 0; x_k = 1; x_l = 2$	1	1	0	0	*T*
$x_i = 0; x_j = 0; x_k > 0; x_l < 2$	1	1	0	0	*T*
$x_i = 0; x_j = 0; x_k = 0$	1	1	1	1	*A*

Note – *i* refers to position of the critical (potentially repeated) letter that appears in all foils. *j* refers to the "interior" position of the latter adjacent to *i* that is repeated in repetition foils and deleted in substitution foils. *k* refers to the other (potentially repeated) position adjacent to *j*. *l* refers to the other "flanker" position adjacent to *k*, and *h* refers to the other "flanker" position adjacent to *i*. In the "fx" ("effects") column: *A* = a guess is made in all conditions; *F* = a guess is made in conditions with repetition in the foil; *f* = a guess is made in conditions without repetition in the foil; *T* = a guess is made in conditions with repetition in the target; *N* = a guess is not made in any condition; *I* = a guess is made only in the condition where both target and foil contain a repetition.

Table D1 delineates the relationship between possible states $\boldsymbol{x}$ and consistency statuses v in each of the four conditions of Experiment 4. Using the categorization of these states in the final column, we can delineate the probability of a guess, which is linearly related to the probability of an error, as follows:

$$P(\text{guess} \mid TRep \,\&\, FRep) = P(A) + P(T) + P(F) + P(I)$$

$$P(\text{guess} \mid TRep \,\&\, FUniq) = P(A) + P(T) + P(f)$$

$$P(\text{guess} \mid TUniq \,\&\, FRep) = P(A) + P(F)$$

$$P(\text{guess} \mid TUniq \,\&\, FUniq) = P(A) + P(f)$$

because A, T, F, f, I, and N are mutually exclusive, exhaustive options.

The predicted simple effects on guessing are therefore:

Target repetition effect when foil is repeated =

$$P(\text{guess} \mid TRep \,\&\, FRep) - P(\text{guess} \mid TUniq \,\&\, FRep) = P(T) + P(I)$$

Target repetition effect when foil is unique =

$$P(\text{guess} \mid TRep \,\&\, FUniq) - P(\text{guess} \mid TUniq \,\&\, FUniq) = P(T)$$

Foil repetition effect when target is repeated =

$$P(\text{guess} \mid TRep \,\&\, FRep) - P(\text{guess} \mid TRep \,\&\, FUniq) = P(F) - P(f) + P(I)$$

Foil repetition effect when target is unique =

$$P(\text{guess} \mid TUniq \,\&\, FRep) - P(\text{guess} \mid TUniq \,\&\, FUniq) = P(F) - P(f)$$

By subtracting either pair of simple effects (the first two or the last two) it can be seen that there is a predicted super-additive interaction of the two forms of repetition of magnitude $P(I)$.

Both foil repetition effects contain the term $P(F) - P(f)$ because in the F rows, the unique foil has the advantage (will not cause a guess) but in the f row, the repeated foil has the advantage. The probabilities of these rows depend on parameters and the stimulus duration. We therefore cannot show that $P(F) > P(f)$ for all possible parameter values, so we cannot state a parameter-free prediction of the model for the effect of foil repetition.

The F rows consist of cases where the E (letter between critical letter and its twin) is perceived without further relevant disambiguation, because this only appears in the repeated target. The f

rows consist of the cases that, in the example in the table, when the initial OE/OB/OL letters are perceived to be adjacent and the subsequent two letters are not perceived at all – OE in the repetition foil is inconsistent with either OB or OL. In other words: these involve the critical letter and its outside flanker being perceived fully and the potential twin and the interior letter not being perceived at all.

Supplementary material

Repeated letters increase the ambiguity of strings: Evidence from identification, priming and same-different tasks

Iliyana V. Trifonova[a,b] and James S. Adelman[a]

[a]University of Warwick, [b]University of York

Results: Experiment 2, Different Trials

Table S1. Mean Reaction Times (ms) and Error Rates (%; in parentheses) by Condition in Experiment 2, Different Trials

	Prime Length							
	7				8			
	Deletion/Control				Identity/Control			
Target Type	Unique		Repeated		Unique		Repeated	
Reference Type	Unique	Repeated	Unique	Repeated	Unique	Repeated	Unique	Repeated
Prime								
Related	870 (36)	784 (19)	815 (22)	682 (5)	876 (35)	772 (22)	807 (23)	691 (5)
Control	878 (34)	782 (17)	809 (22)	694 (6)	883 (33)	761 (20)	802 (23)	676 (6)

Results: Experiment 4a

Same Trials

The effect of target type was significant, $\chi^2(1) = 4.554$, $p = .033$.

The effect of presentation was not significant, $\chi^2(1) = 1.509$, $p = .219$, nor was the interaction between the two factors, $\chi^2(1) < 1$.

Table S2. Mean Accuracy (%) by Condition in Same Trials, Experiment 4a

	Presentation	
Target type	Normal	Enhanced
Repeated	78.8	79.6
Unique	75.7	77.0

Different Trials

The effect of target type was significant, $\chi^2(1) = 33.178$, $p < .001$.

The effect of foil type was highly significant, $\chi^2(1) = 148.391$, $p < .001$.

The interaction between target type and foil type was significant, $\chi^2(1) = 5.691$, $p = .017$, as was the interaction between presentation and foil type, $\chi^2(1) = 6.63$, $p = .01$.

The three-way interaction between foil type, target type and presentation was marginally significant, $\chi^2(1) = 3.012$, $p = .083$.

Table S3. Mean Accuracy (%) by Condition in Different Trials, Experiment 4a

	Foil Type			
	Repeated (Wrong repetition)		Unique (Substitution)	
	Presentation			
Target type	Normal	Enhanced	Normal	Enhanced
Repeated	32.6	29.6	56.8	63.4
Unique	43.0	41.8	64.5	64.8

Results: Experiment 4b

Same Trials

The effect of target type was not significant, $\chi^2(1) = 2.141$, $p = .143$. Although there was a trend for higher accuracy in the repeated target type condition, it was less pronounced than in Experiment 4a.

The effect of presentation type was approaching significance, $\chi^2(1) = 3.002$, $p = .083$. The interaction between the two factors was not significant, $\chi^2(1) < 1$.

Table S4. Mean Accuracy (%) by Condition in Same Trials, Experiment 4b

	Presentation	
Target type	Normal	Enhanced
Repeated	77.5	79.3
Unique	75.8	77.9

Different Trials

Consistent with the results of Experiment 4a, the effects of target type and foil type were significant, $\chi^2(1) = 35.684$, $p < .001$, $\chi^2(1) = 239.361$, $p < .001$, as was their interaction $\chi^2(1) = 8.675$, $p = .003$.

The interaction between foil type and presentation was, however, not significant, $\chi^2(1) <1$, nor was the three-way interaction between foil type, target type and presentation type, $\chi^2(1) <1$.

Table S5. Mean Accuracy (%) by Condition in Different Trials, Experiment 4b

	Foil Type			
	Repeated		Unique	
	Presentation			
Target type	Normal	Enhanced	Normal	Enhanced
Repeated	29.4	29.9	64.2	65.1
Unique	42.8	41.7	68.5	67.7