Behav Res (2014) 46:1052–1067
DOI 10.3758/s13428-013-0442-y

A behavioral database for masked form priming

James S. Adelman · Rebecca L. Johnson · Samantha F. McCormick · Meredith McKague · Sachiko Kinoshita · Jeffrey S. Bowers · Jason R. Perry · Stephen J. Lupker · Kenneth I. Forster · Michael J. Cortese · Michele Scaltritti · Andrew J. Aschenbrenner · Jennifer H. Coane · Laurence White · Melvin J. Yap · Chris Davis · Jeesun Kim · Colin J. Davis

Published online: 1 February 2014
© Psychonomic Society, Inc. 2014

Abstract Reading involves a process of matching an orthographic input with stored representations in lexical memory. The masked priming paradigm has become a standard tool for investigating this process. Use of existing results from this paradigm can be limited by the precision of the data and the need for cross-experiment comparisons that lack normal experimental controls. Here, we present a single, large, high-precision, multicondition experiment to address these problems. Over 1,000 participants from 14 sites responded to 840 trials involving 28 different types of orthographically related primes (e.g., *castfe*–CASTLE) in a lexical decision task, as well as completing measures of spelling and vocabulary. The data were indeed highly sensitive to differences between conditions: After correction for multiple comparisons, prime type condition differences of 2.90 ms and above reached significance at the 5% level. This article presents the method of data collection and preliminary findings from these data, which included replications of the most widely agreed-upon differences between prime types, further evidence for systematic individual differences in susceptibility to priming, and new evidence regarding lexical properties associated with a target word's susceptibility to priming. These analyses will form a basis for the use of these data in quantitative model fitting and evaluation and for future exploration of these data that will inform and motivate new experiments.

Keywords Visual word recognition · Lexical decision · Orthographic priming · Megastudies

Introduction

The everyday activity of reading involves correctly selecting from one's vocabulary the viewed word from among a variety of candidate words with some or many of the same features.

J. S. Adelman (✉)
Department of Psychology, University of Warwick, Coventry CV4 7AL, UK
e-mail: J.S.Adelman@warwick.ac.uk

R. L. Johnson
Skidmore College, Saratoga Springs, USA

S. F. McCormick · C. J. Davis
Royal Holloway, University of London, Egham, UK

M. McKague
University of Melbourne, Parkville Vic, Australia

S. Kinoshita
Macquarie University, Sydney, Australia

J. S. Bowers · C. J. Davis
University of Bristol, Bristol, UK

J. R. Perry · S. J. Lupker
University of Western Ontario, London, Canada

K. I. Forster
University of Arizona, Phoenix, USA

M. J. Cortese
University of Nebraska, Omaha, USA

M. Scaltritti · A. J. Aschenbrenner
Washington University, St. Louis, USA

J. H. Coane
Colby College, Waterville, USA

L. White
Plymouth University, Plymouth, UK

M. J. Yap
National University of Singapore, Singapore, Singapore

C. Davis · J. Kim
Marcs Institute, University of Western Sydney, Bankstown, Australia

How the relevant candidates are evaluated in terms of matches and mismatches in the identity and ordering of letters is a major current concern in visual word recognition research. This concern is reflected in explicit computational models (e.g., Adelman, 2011; Davis, 2010; Norris & Kinoshita, 2012), as well as a wealth of experimental research, much of it using the masked form priming paradigm developed by Forster and Davis (1984) (Forster, Davis, Schoknect, & Carter, 1987). In these experiments (e.g., Davis & Bowers, 2006; Davis & Lupker, 2006; Grainger, Granier, Farioli, Van Assche, & van Heuven, 2006; Perea & Lupker, 2003), the presentation of a target stimulus for a lexical decision is preceded by a brief presentation of a potentially related (nonword) prime stimulus. From the extent to which responses to a word target are faster following a related nonword prime than following an unrelated prime, researchers make inferences regarding similarity between the processing evoked by the related prime and the processing evoked by a veridical presentation of the target.

Many of these experiments have provided indications of important qualitative differences among different types of primes, allowing researchers to distinguish among classes of models. However, as models have become more sophisticated in light of these data, evaluating them by a short list of qualitative criteria alone has become problematic. More than one model may produce the correct qualitative pattern, and modelers have begun to use more quantitative criteria, such as the correlation between the observed priming and priming predicted by a given model (e.g., Adelman, 2011; Davis, 2010). However, calculating such correlations on the basis of the combination of priming effects from several different experiments has key drawbacks.

First, combining data across experiments lacks the kinds of experimental controls we would normally expect in our studies. We would ordinarily aim to ensure that our estimates of priming of different kinds come from the same population of participants and, ideally, from a within-subjects experiment. Moreover, we would ordinarily use the same targets for each type of priming of interest, in order to avoid contamination by moderating properties of the target words. Furthermore, we would not use different equipment, font sizes, and so forth in investigating different types of priming.

Second, the size of many of the priming effects has not been estimated with great precision, when this was not needed for the comparison of interest in the original study. Typically, a priming effect (or difference between two priming effects) does not reach significance if it is not at least 10 ms (even ignoring the issue of multiple comparisons), implying a 95% confidence interval on the size of the effect with a range of 20 ms or so. Given that the range of priming effects obtained in lexical decision with nonword primes and word targets—the usual paradigm of interest—is only around 50 ms, this level of precision can often be insufficient for the purposes of assessing quantitative model predictions. Contemporary models, such as letters in time and retinotopic space (LTRS; Adelman, 2011), the spatial coding model (SCM; Davis, 2010), and the Bayesian reader (Norris & Kinoshita, 2010), readily make predictions of differences between conditions of less than 10 ms, and this is of particular interest, for instance, in the use of derangements (permutations of stimulus order that leave no letter in its original position; see, e.g., Guerrera & Forster, 2008; Lupker & Davis, 2009), where nonsignificant results of this magnitude have been observed. Moreover, these models can make predictions that are quite similar to one another; in Adelman's comparison of the LTRS and SCM models, the average absolute discrepancy in predicted priming between the two models was 8 ms.

Third, the estimates of the size of the priming effects are biased upward by the processes involved in selecting experiments for publication. An experiment that has no significant effects is highly unlikely to be published, so an experiment in which the noise in the data from the control condition happens to make it unusually slow is more likely to be published than one where it is unusually fast, which is likely to reach the file drawer. Thus, published priming effects will be, on average, larger than the true effects.

Fourth, relatively few of the effects shown in these studies have been subject to published replication, which is a concern with a false positive rate of 1 in 20, regardless of other concerns (see, for instance, the special issue edited by Pashler and Wagenmakers, 2012).

Some elements of the mega-study approach (e.g., Balota et al., 2007; for a review, see Balota, Yap, Hutchison, & Cortese, 2012) can appropriately deal with these concerns. First, as a single large study with within-subjects manipulations, a mega-study creates no systematic relationship between lexical predictor variables and properties of the participants. Second, with more participants and stimuli, a mega-study offers greater precision of estimates of any effects (provided that the increase in quantity of data is sufficient to countermand other sources of variability that might be introduced, such as site differences). Third, with no particular comparison reaching significance being needed for the study to be of interest, mega-studies do not get stuck in the file drawer. Fourth, mega-studies offer (conceptual) replications that might not otherwise occur for many effects. However, due to the nature of priming manipulations—the presence of two stimuli on a trial, the need for a baseline condition, and the fact that most pairs of words are unrelated or weakly related—a priming mega-study requires a more controlled approach than do other paradigms where an exhaustive selection (of a subset of items with some property) would be possible and a random selection from among these items would be useful; that is, the various related and unrelated conditions in a priming study must be selected a priori (for this approach to semantic/associative priming, see Hutchison, Balota, Cortese, & Watson, 2007; Hutchison et al. 2014).

The present study

With these points in mind, the present study was designed to produce a large masked priming data set that could serve as the basis for a wide range of analyses that would be useful in assessing models. To do so, researchers from 14 different universities collected data using 28 prime types with 420 word targets (and 420 nonword foils). This approach differs from most mega-studies in two major respects: The initial focus of the study is the different conditions designed into the experiment, rather than lexical properties of words sampled across their natural distribution, and the comparisons of these conditions are controlled by counterbalancing rather than by covarying out potential confounders. In these respects, this study is more like a version of an ordinary experiment that is enlarged in terms of words and participants than like other mega-studies.

Prime conditions

Our choices of prime conditions reflect several theoretical and empirical motivations. Our goal was to produce a general database of different types of potentially theoretically relevant primes, rather than to produce yet another experiment that purported to decide between two contemporary theories. We adopted a large range of conditions that models should account for, because (1) doing so will extend the utility of the data set to future, not just contemporary, models of orthographic processing in word recognition; (2) on the sheer balance of probabilities, some of the past experimental comparisons almost certainly produced a wrong result (of type I or type II type); and (3) having a large number of conditions is the most constraining approach when models depend on numerical parameters in a complex way. Conditions where all contemporary models agree on the direction of the qualitative effects are still important to include in a data set of this nature, because they avoid the possibility that modelers will be able to invoke variations in these numerical parameters across different experiments to accommodate patterns that would otherwise be incompatible in their models (see Adelman & Brown, 2008a, for further discussion). There are several sets of conditions for which models like LTRS can predict several different orderings, depending on the parameters controlling different processing speeds. However, not all of the orderings that LTRS can predict will come about if these speed parameters have to take values that work on the uncontroversial effects. Moreover, since yet unexplored interactions between noncontroversial effects and lexical properties of the targets of priming might be informative, opening up the possibility of such exploratory analysis is one of the major motivations for the collection of mega-study data.

One important issue was to select prime conditions having variants that differed in position—initial, medial or final—in order to address claims regarding the relative importance of exterior and interior letters, arising primarily from other paradigms (e.g., Humphreys, Evett, & Quinlan, 1990).

We also considered it important to include primes created by various amounts of insertion (e.g., *pragkise*–PRAISE) and deletion (e.g., *prse*–PRAISE; also known as superset and subset primes, respectively; see, e.g., Grainger et al., 2006; Van Assche & Grainger, 2006). These conditions provide evidence regarding the relative importance of different positions, the flexibility of positional representation, and the balance of positive and negative evidence in lexical matching.

We further considered it important to include transposition (e.g., *priase*–PRAISE) and substitution (e.g., *prnvce*–PRAISE) primes involving various positions. The evidence that transpositions produce more priming than do corresponding substitutions (e.g., Perea & Lupker, 2003) suggests that letter identity and position are encoded separately (i.e., slot-based coding of letter position [e.g., McClelland & Rumelhart, 1981] is inconsistent with the data) and is a key motivation for the recent development of new models of letter identification and lexical matching. We also included neighbor-once-removed primes (e.g., *prihse*–PRAISE) that combine transposition with substitution of a transposed letter. Evidence that this condition produces less priming than the substitution alone (Davis & Bowers, 2006) is inconsistent with coding schemes that base matching on open bigrams (i.e., representations of letter pair orders where letter pairs need not be adjacent) alone (e.g., Grainger & van Heuven, 2003).

We also considered more extreme transposition primes (involving many changes in letter order—e.g., *rpiaes*–PRAISE), because the absence of significant priming in such conditions (e.g., Guerrera & Forster, 2008; Lupker & Davis, 2009) has been taken as informative regarding forms of inhibition that might operate in addition to the facilitation from the overlap between prime and target.

Our two unrelated conditions—a wordlike pseudoword condition and a more arbitrary letter string condition—represent two strategies for selecting baselines from which to calculate priming effects. In contrast to these lower limits, the identity condition represents a presumed upper limit on the amount of priming within the paradigm.

Individual differences

With the large array of words in this study, there is considerable variability in their lexical properties, and these properties could operate as moderators of the priming effects. Those moderators would be open to analysis by regression of this study's data. Such use is a further motivation for this kind of study. The details of such moderating effects can act as additional constraints on models of the relevant processes. Furthermore, we took the opportunity to collect brief measures of individual differences in spelling and vocabulary—which are moderators of priming effects (Andrews & Hersch, 2010; Andrews & Lo, 2012)—to extend the possibilities for additional uses of the data set.

Method

Participants were recruited at 14 different university sites for course credit or monetary compensation in multiples of 28, according to the counterbalancing scheme described below. Problems with timing responses with the equipment at the University of Nebraska, Omaha, led to those participants being excluded from further analysis. Participants whose accuracy at primed lexical decision was below 75% were replaced.[1] Table 1 summarizes the number of participants, including the number excluded for errors or as excess to counterbalancing, and the form of compensation at each site. Since our goal was to produce as precise an estimate of each priming effect as possible, each site was asked to provide as many participants as possible within a fixed time window.

Design

Lexical decision times were measured in response to words primed with 28 types of primes, detailed in Table 2. Prime type was varied within subjects, with participants within each site spread evenly over 28 counterbalancing lists. All lists contained all targets, and targets were paired with a different prime type in each list, with the constraint that each list contained an equal number (15) of each prime type.

Apparatus and software

Visual presentation apparatus varied with site, as detailed in Table 1, but all sites used a 60-Hz refresh rate setting. Moreover, an estimate of the typical viewing distance (since no chinrest or other control on head position was used) was used to modify the scripts for the DMDX stimulus presentation software (Forster & Forster, 2003) to adjust the size of the stimuli so that the width of each character in a target was approximately 1° of visual angle.

Some sites used button boxes for responses to two-alternative choice tasks; the remainder used keyboard presses (left and right shift-keys). Although keyboards are not particularly precise devices for measuring individual response times (RTs), Ulrich and Giray (1989) showed that such problems can only have their influence by increasing variance, reducing power, which can be counteracted by increasing sample size, and Damian (2010) showed that the additional variance introduced by keyboards is negligible, relative to human variability for numbers of trials orders of magnitude fewer than in the present mega-study. All sites used numerical keyboard presses as responses for the vocabulary task with four alternatives. The monitor and button box or keyboard used are listed in Table 1.

These were the only differences in the manner in which the experiment was presented at the various sites.

Stimuli

Four hundred twenty word stimuli, all six letters long, were selected to be targets in the primed lexical decision task, and for each of these, 27 nonwords were chosen to be primes, 1 for each of the nonidentity prime types detailed in Table 2 (where lexicality was determined by reference to CELEX; Baayen, Piepenbrock, & Gulikers, 1995). All word targets had a frequency above zero in CELEX, HAL (Burgess, 1998), and SUBTLEX (Brysbaert & New, 2009) and had lexical decision accuracy of at least 80% in the English Lexicon Project (Balota et al., 2007). No target was listed in CELEX as having a derived or inflected morphemic structure. One hundred twenty targets had at least one higher frequency orthographic neighbor. Other details of frequency and the orthographic and phonological characteristics of the word targets are given in Table 3.

No target contained the same letter twice, nor did any prime other than the repeated-insertion (123DD456) prime. Some prime type conditions could be created in more than one way (e.g., since we used six-letter targets, a single medial substitution could occur in four positions), as indicated in the table; across targets, these subconditions were used equally often. Inserted or substituted letters were chosen at random without replacement from those not in the target.

Two of the prime types were designed to be orthographically unrelated baselines. The arbitrary baseline was composed of six letters not in the target, pseudorandomly chosen without replacement, with the constraint that the prime be a nonword. The pseudoword primes were created to be wordlike (and likely pronounceable) in the same way as the nonword foils, which we will now describe.

Four hundred twenty nonword foils were constructed using an algorithm that pseudorandomly replaced two letters of a real word in such a way that the resulting string contained no repeated letters and each of its trigram frequencies exceeded a minimum value of one per million (based on the CELEX database). The real-word inputs to this algorithm were the 420 word targets. In this way, it was ensured that the nonword

[1] For this purpose, failure to respond before a 2,000-ms timeout, described in the Procedure section (0.47% of all trials), was counted as an error. Data are included in the downloadable database for all participants who were excluded or replaced for the analyses we present here.

Table 1 Number of participants from each contributing site

	Subjects Tested	Subjects Tested / 28	Excluded Due to Error or Equipment	Excess to Counterbalancing	Subjects Used	Complete Sets of 28 Used	Compensation	Monitor	RT Device
RHUL	217	7.75	5	16	196	7	Cash or CC	Samsung 793DF CRT	Custom DMDX PIO-12 BB
Skidmore	197	7.04	0	1	196	7	Cash or CC	Dell 1907FPVt LCD	Dell L100 KB
Warwick	119	4.25	2	5	112	4	CC	Sony CPD-G200 CRT	NMB RT2158TWUK KB
Melbourne	66	2.36	1	9	56	2	CC	Sony E230 CRT	Custom DMDX PIO-12 BB
Macquarie	65	2.32	3	6	56	2	CC	NEC 4FG Multisync CRT	Custom DMDX PIO-12 BB
UWO	60	2.14	4	0	56	2	Cash or CC	LG Flatron W2242TQ LCD	HP KB38211 KB
Bristol	59	2.11	0	3	56	2	CC	AOC F1770 CRT	Viglen KU-0325 KB
WUSTL	57	2.04	1	0	56	2	Cash or CC	iMac MA590LL integral LCD	Apple KY6310LVLVZSA KB
Arizona	31	1.11	3	0	28	1	CC	NEC 14" Multisync C500 CRT	Custom DMDX PIO-12 BB
MARCS	31	1.11	3	0	28	1	Cash or CC	Mitsubishi DV997FDB CRT	Custom DMDX PIO-12 BB
Nebraska	29	1.04	29	0	0	0	CC	Dell 1908FPt LCD	Dell U 4739 KB
Colby	28	1.00	0	0	28	1	Cash or CC	Dell P2210t LCD	Dell SK-8115 KB
Plymouth	28	1.00	0	0	28	1	CC	Hanns-G AG172D LCD	Viglen KB-0325 KB
Singapore	28	1.00	0	0	28	1	Cash	Viewsonic E72f+ CRT	Dell SK-8110 KB

Note. MARCS = Marcs Institute, University of Western Sydney; RHUL = Royal Holloway, University of London; UWO = University of Western Ontario; WUSTL = Washington University, St. Louis; CC = course credit; KB = keyboard; BB = button box.

Table 2 Prime types forming conditions of the experiment

Prime Type	Code Relative to 123456	Abbreviation	e.g.: DESIGN
Identity	123456	ID	design
Initial transposition	213456	TL12	edsign
Medial transposition	132456/124356/ 123546	TL-M	desgin
Final transposition	123465	TL56	desing
2-apart transposition	143256/125436	NATL-24/35	degisn
3-apart transposition	153426	NATL25	dgsien
Medial deletion	13456/12456/12356/ 12346	DL-1M	dsign
Final deletion	12345	DL-1F	desig
Central double deletion	1256	DL-2M	degn
All-transposed	214365	T-All	edisng
Transposed halves	456123	TH	igndes
Half	123/456	SUB3	des
Reversed halves	321654	RH	sedngi
Interleaved halves	415263	IH	idgens
Reversed-except-initial	165432	RF	dngise
Initial substitution	d23456	SN-I	pesign
Medial substitution	1d3456/12d456/ 123d56/1234d6	SN-M	desihn
Final substitution	12345d	SN-F	desigj
Neighbor-once-removed	12d356/13d456/ 124d56/123d46	N1R	dslign
Central double substitution	12dd56	DSN-M	dewvgn
Central insertion	123d456	IL-1M	desrign
Central double insertion	123dd456	IL-2M	desaxign
As above, repeated letter	123DD456	IL-2MR	deshhign
Central quadruple substitution	1dddd6	EL	dzbtkn
Prefix	d123456	IL-1I	mdesign
Suffix	123456d	IL-1F	designl
Unrelated pseudoword	dddddd	ALD-PW	voctal
Unrelated arbitrary	dddddd	ALD-ARB	cbhaux

Note. Where multiple codes are indicated, equal numbers of targets participate in each of these subconditions. Where d or D is indicated, a random letter not present in the target is used; where d is indicated more than once, the same letter is not reused; where D is indicated more than once, the same letter is reused.

foils were well-formed English stimuli with orthographic structures that closely matched those of the word targets. The selection was constrained such that none of the primes for the nonwords (constructed analogously to the primes for the words) was a word; this constraint—and the prohibition on repeated letters—necessitated a custom program. Orthographic characteristics of the nonword foils are given in Table 3.

To assess spelling ability, a set of items based on inconsistent and unusual sound–spelling correspondences was used, consisting of 42 words (e.g., ELEMENTARY) and 40 nonwords that were words modified to contain typical spelling errors (e.g., REFRENCES, BENAFIT, TOUNGE), based on a list from Burt and Tate (2002).

The Shipley (1940) vocabulary test has 40 target words of increasing difficulty (from TALK to PRISTINE), each associated with one correct synonym and three foils.

Procedure

Participants first completed the primed lexical decision task. On each trial, a 300-ms initial presentation of a central fixation cross (+) was followed by a 200-ms blank display, after which a hash (##########) mask was presented for 500 ms. Then the prime was displayed in lowercase at five-eighths size for 50 ms, before the target appeared in uppercase until either the participants responded with a left or right response for a nonword or word, respectively, or 2,000 ms had elapsed. If an incorrect response was given or 2,000 ms elapsed without a response, corrective feedback was given. Instructions preceding this task described the sequence of events, omitting mention of the prime, and indicated the timed nature of the task, while indicating that accuracy should not be unduly sacrificed.

The procedure was similar for the spelling items that followed in a new block; — that is, the task was lexical decision—but without the prime and 2,000-ms cutoff. Participants were given new instructions, that these items were chosen to be difficult to spell, and accuracy was emphasized.

Finally, a computerized version of the vocabulary portion of the Shipley (1940) test was administered, displaying the target above four numbered potential synonyms in turn for each of the 40 items, for a numerical key response.

Results

All the trial-by-trial data, including those for participants excluded from analysis here, are available for download from https://files.warwick.ac.uk/jadelman2/browse#FPP. The data are available as text files or an Excel spreadsheet, including details of the excluded trials and the calculation of the condition means.

Overall priming results

Trials with associated RTs of more than 1,500 or less than 150 ms (0.77% and 0.06% of correct trials, respectively, as illustrated in Fig. 1) were discarded—comparably with prior studies—for the purposes of the following presentation of the data (again, "excluded" trials are included in the full database available to the database user). Table 4 presents the mean

Table 3 Properties of lexical decision targets, according to Elexicon Web interface

Orthography						
	Words			Nonword foils		
	Min.	Max.	Mean	Min.	Max.	Mean
Orthographic *N*	0	8	1.29	0	6	0.61
Bigram frequency	442.8	4,187.6	1,798.11	390.4	3,937.4	1,732.14
OLD20	1.45	3.00	2.12	1.60	2.95	2.29
Frequency (*words*)						
	Min.	Max.	Mean	Mean (log.)		
HAL frequency	62	283,001	13,258	8.08		
SUBTLEX freq.	0.12	501.33	18.73	2.33		
SUBTLEX CD	0.07	85.35	5.39	2.16		
Phonology (*words*)						
	Min.	Max.	Mean			
Phonological *N*	0	35	2.96			
Phonographic *N*	0	6	0.71			
PLD20	1.00	4.00	2.01			
No. of phonemes	3	7	5.09			
No. of syllables	1	3	1.82			

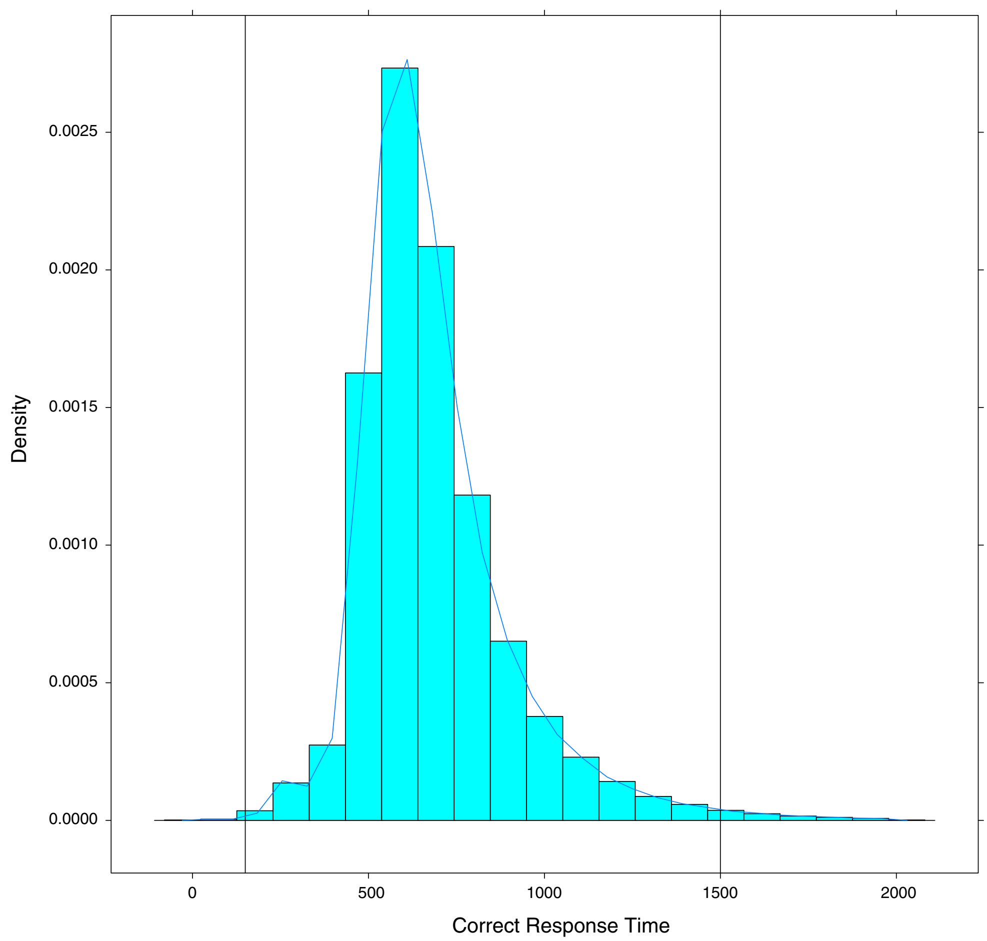

Fig. 1 Histogram of all correct response times within the 2-s timeout

Table 4 Mean correct response times (RTs) for each prime type for word targets

	Code	RT	Priming-ARB	Priming-PW
ID	123456	634.48	42.69	37.89
DL-1F	12345	642.93	34.23	29.44
IL-1F	123456d	643.51	33.66	28.86
TL56	123465	644.70	32.46	27.67
TL-M	132456/124356/ 123546	645.74	31.42	26.62
DL-1M	13456/12456/ 12356/12346	647.60	29.56	24.77
SN-F	12345d	647.71	29.45	24.66
SN-I	d23456	648.00	29.16	24.37
TL12	213456	648.13	29.03	24.23
IL-1M	123d456	648.16	29.00	24.21
IL-1I	d123456	650.49	26.67	21.88
SUB3	123/456	651.34	25.83	21.03
IL-2MR	123DD456	651.68	25.48	20.69
DL-2M	1256	652.25	24.91	20.12
SN-M	1d3456/12d456/ 123d56/1234d6	654.48	22.68	17.88
N1R	12d356/13d456/ 124d56/123d46	655.40	21.77	16.97
NATL-24/35	143256/125436	656.97	20.20	15.40
IL-2M	123dd456	657.74	19.42	14.63
T-All	214365	660.39	16.77	11.98
DSN-M	12dd56	662.23	14.94	10.14
RH	321654	663.73	13.44	8.64
NATL25	153426	667.25	9.91	5.11
IH	415263	668.26	8.90	4.11
TH	456123	668.36	8.80	4.01
ALD-PW	dddddd	672.37	4.80	0.00
RF	165432	674.30	2.86	−1.94
EL	1dddd6	674.82	2.34	−2.46
ALD-ARB	dddddd	677.17	0.00	−4.80

correct RTs to words for each of the 28 prime types and the resultant priming estimates against each unrelated baseline.

Although the size of the data set (and design matrix for the analysis) is prohibitive for ANOVA or other linear model analysis using most modern computational software,[2] we were able to use direct computation of the sums of squares (literally adding up the squares of residuals for particular models, with model fitting based on marginal means, which is not how modern software typically computes the values to appear in an ANOVA table) in combination with Clark's (1973) pseudo-*F* calculations to produce a reasonable estimate of the mean-squared error (Clark's Equation 14) in the estimation of the comparisons between conditions, taking into account random effects associated with both subjects and items. Applying Tukey's HSD procedure to adjust for multiple comparisons with this estimate of the mean-squared error implies that differences between conditions of 2.90 ms at the 5% level, 3.24 ms at the 1% level, 3.37 ms at the 0.5% level, and 3.66 ms at the 0.1% level are significant.

Comparisons of priming by site

Figure 2 illustrates the quality of agreement between the priming estimates for each testing site. Even the sites with fewer participants correlate very highly with the average of all sites, although (as one would expect) not so well with each other. However, this fact does not rule out differences in the magnitude of priming (e.g., a site with half the priming effect would show a perfect correlation), which are suggested by the varying slopes of the regression lines.[3] Nevertheless, the agreement between sites is notably good, considering the variation in equipment used at different sites. We also specifically compared the RTs from each priming condition collected with LCD monitors against those collected with CRT monitors. Although RTs from the LCD sites were 41.47 ms longer than those from the CRT sites, paired $t(27) = 68.96$, the correlations of RTs from each priming condition across the two monitor types was $r(28) = .965$, and the correlation of priming effects relative to the arbitrary baseline was $r(27) = .960$.

Spelling, vocabulary, and site differences

The mean spelling score was 77.7% ($SD = 8.8\%$). The mean vocabulary score was 74.3% ($SD = 9.7\%$). Spelling and vocabulary were well correlated across participants, $r(924) = .422$, $p < 10^{-15}$.

The mean for each site in spelling, vocabulary, baseline RT, and three major priming effects is given in Table 5. Substantial variability is shown in the estimates of priming at this level, despite the good correlations among sites.

Correlations of the spelling and vocabulary variables with each of the condition mean RTs and the priming effects are presented in Table 6, as are correlations of the sum and difference of the spelling and vocabulary scores and their *z*-scores[4] with RTs and priming effects. There is a clear (expected) pattern such that responses are faster for those with

[2] The forms of analysis that could not be performed due to the computational memory requirements are those whose computation includes the calculation of the pseudoinverse of the design matrix. In modern software, such as SPSS, SAS, and R, this is routinely used as part of the fitting of linear models, including mixed effects models.

[3] The Deming regression in those lines corrects for attenuation or regression dilution due to noise in the *x*-observations insofar as its ratio with that in the *y*-observations can be predicted from sample size.

[4] With two positively correlated variables, the standard (equal-variance) principal components analysis gives the sum and difference of the *z*-scores, divided by $\sqrt{2}$, so for the purposes of correlation these are equivalent.

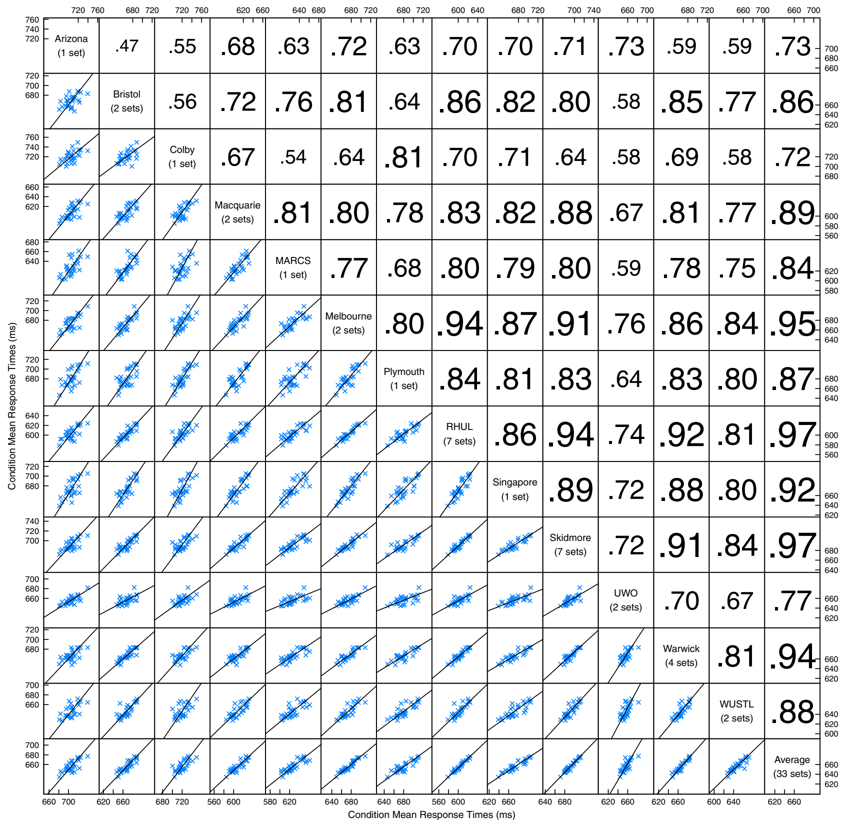

Fig. 2 Comparison of priming effects at the different sites using condition mean correct response times at each site. Above/right of diagonal: Correlation coefficient between condition means at each pair of sites and (weighted) average of sites. Below/left of diagonal: Scatterplot of these condition means with Deming regression; this technique differs from ordinary linear regression in that it allows for noise in the *x*-observations of known size, relative to the *y*-observations (rather than assuming that there is no noise in the *x*-observations). Here, this ratio was set to that implied if differences were caused only by sample size

Table 5 Mean accuracy on spelling and vocabulary trials (%) for each contributing site, with baseline response time (in milliseconds) from unrelated prime trials and priming effects (in milliseconds) for identity, medial one-letter-different and medial transposed-letter primes

Site (# of sets)	Accuracy (primed LDT)	Spelling	Vocabulary	Baseline ARB	Baseline PW	ID Priming (-ARB)	1LD Priming (-ARB)	TL Priming (-ARB)	ID Priming (-PW)	1LD Priming (-PW)	TL Priming (-PW)
Arizona (1 set)	89.07	71.17	64.55	739.44	719.20	58.34	42.41	55.82	38.10	22.18	35.58
UWO (2 sets)	90.05	75.70	69.73	682.20	666.06	38.01	34.03	36.51	21.88	17.89	20.37
Bristol (2 sets)	90.85	77.07	73.53	683.13	688.54	32.35	19.96	15.67	37.76	25.37	21.07
Macquarie (2 sets)	91.37	75.94	74.42	625.00	614.44	40.25	24.51	34.31	29.69	13.95	23.75
Melbourne (2 sets)	91.55	79.36	73.88	708.64	694.74	59.07	47.12	48.57	45.17	33.22	34.67
Warwick (4 sets)	92.35	77.47	71.36	683.36	682.20	34.08	17.79	19.62	32.92	16.63	18.47
RHUL (7 sets)	92.36	77.96	73.71	624.24	623.86	45.43	30.48	32.58	45.04	30.09	32.20
Plymouth (1 set)	92.52	74.48	72.86	711.30	695.47	64.65	45.90	32.45	48.82	30.07	16.62
MARCS (1 set)	92.79	79.75	77.95	649.53	642.04	43.77	27.86	36.17	36.27	20.37	28.67
Singapore (1 set)	93.26	82.10	76.07	705.33	693.17	66.80	37.50	48.98	54.64	25.35	36.83
Skidmore (7 sets)	93.29	77.56	76.11	710.69	707.94	39.32	25.19	31.73	36.57	22.44	28.97
WUSTL (2 sets)	94.02	81.34	81.56	664.54	664.41	35.59	29.45	23.48	35.46	29.33	23.35
Colby (1 set)	94.89	79.14	80.18	732.19	724.39	27.94	26.42	17.57	20.14	18.62	9.77
Average (33 sets)	92.34	77.70	74.30	677.17	672.37	42.69	22.68	31.42	37.89	17.88	26.62
Highest - lowest	5.82	10.93	17.01	115.20	109.05	38.86	29.33	40.15	34.50	19.27	27.06

better spelling and vocabulary (e.g., Yap, Balota, Sibley, & Ratcliff, 2012). Furthermore, those with better spelling and vocabulary showed less priming.

Turning to the difference between spelling and vocabulary, spelling had a stronger relationship with overall RTs. As a consequence, the difference scores (SpellMinusVocab) also correlated negatively with RTs. On priming, however, if anything, the effect was in the reverse direction. In the final column of Table 6, there are 20 out of 27 (sign test: p =.019) positive correlations between ZSpellMinusZVocab and priming; these results indicate greater priming for those participants whose spelling was relatively better than their vocabulary. However, this pattern was weak, and none of the correlations was significant in its own right.

Item-level analysis of targets

We examined whether targets with particular lexical properties were particularly susceptible to or immune from priming by correlating lexical properties of the word targets with the priming as measured by subtracting the mean of all the related conditions from the mean of the two control conditions. We calculated the split-half reliability of this measure with 100 splits of the participants (with one "half" having 17 participants from each counterbalancing, and the other 16); the average was .116. In addition to log. word frequency taken from CELEX (Baayen et al., 1995), the following lexical properties were taken from the English Lexicon Project(Balota et al., 2007): log. HAL frequency, log. SUBTLEX frequency, log. SUBTLEX contextual diversity, mean bigram frequency, number of homophones, number of syllables, number of phonemes, orthographic neighborhood size, phonological neighborhood size, phonographic neighborhood size, and the Levenshtein-based neighborhood variables proposed by Yarkoni, Balota, and Yap (2008), as detailed in Table 7. The Levenshtein variables are calculated as the average orthographic or phonological Levenshtein distance of the 20 nearest words of a given target word; these are known as OLD20 (orthographic Levenshtein distance 20) and PLD20 (phonological Levenshtein distance 20). The zero-order correlations in that table show that PLD20 had the numerically strongest relationship with priming. All variables except homophony had at least a marginal relationship between all these variables and priming. The correlations among the various predictors of priming indicate that the influences of these variables on priming might not be unique.

We addressed the uniqueness of the effects of each of these variables in 16 multiple regression analyses that combined each of the four frequency (or contextual diversity) counts with each of four strategies for including neighborhood variables: first, including them all; second, including all three (N) measures based on one-letter and one-phoneme different neighbors; third, including only the orthographic and phonological neighborhood size, excluding the phonographic neighborhood variable; and fourth, using only the Levenshtein-based OLD20 and PLD20 measures. These analyses are summarized in Table 8.

In analyses where PLD20 was included, it was a significant predictor of priming and was the only neighborhood variable that predicted priming (at the 5% level). When PLD20 was excluded, orthographic N did usually predict priming. The only other variables to reach significance were those based on the subtitle-based corpus. Log. subtitle frequency predicted priming in all analyses in which it appeared, whereas log. subtitle contextual diversity was significant only when PLD20 was absent from the regression.

Discussion

We have presented the first large single-experiment database of masked form priming data in order to address four main limitations of using ad hoc databases composed of several experiments. First, a lack of control exists when experiments are combined. Although here we used multiple sites with differing equipment, such differences were spread uniformly across conditions, rather than confounded with condition. Second, there has been a lack of precision (and power) in priming estimates from prior studies. In the present study, differences between conditions of around 3 ms could be considered significant, even with correction for (378) multiple comparisons. Examination of subsets of the data showed the preceding concerns to be valid: Although the data set as a whole had good reliability and sites correlated well with one another, there were still substantial differences in estimates of individual priming effects across different sites, especially for those with fewer participants. Third, publication bias might overestimate priming effects, and fourth, past studies had not been subject to replication. We now compare our results with those of some previous studies.

Exterior versus interior letters

Examination of conditions in which the difference between the prime and target was the substitution of a single letter provided surprising evidence for the importance of central letters. Medial substitutions (SN-M: *desihn*–DESIGN) produced less priming than did initial (SN-I: *pesign*–DESIGN) and final (SN-F: *desigj*–DESIGN) substitutions, with the two ends not differing from one another. This is consistent with the finding of Perea and Lupker (2003) that final, but not medial, double substitutions produce priming with five-letter words. However, this result contrasts with the finding of Schoonbaert and Grainger (2004) in French that double substitutions produced priming only when the substitution involved the last

Table 6 Correlations between individual difference variables—spelling, vocabulary, their sum, and their difference—and condition response times and priming effects (against the arbitrary unrelated prime baseline) at the subject level

Correlation with:	RT (ms)					Priming (ms)						
Prime Type	Spelling	Vocabulary	SpellPlus Vocab	SpellMinusVocab	ZSpellPlusZVocab	ZSpellMinus ZVocab	Spelling	Vocabulary	SpellPlus Vocab	SpellMinus Vocab	ZSpellPlusZ Vocab	ZSpellMinus ZVocab
ALD-ARB	−.231	−.156	−.228	−.053	−.230	−.070						
ALD-PW	−.222	−.135	−.210	−.066	−.212	−.081	−.027	−.042	−.041	.016	−.041	.013
DL-1F	−.235	−.135	−.217	−.077	−.219	−.093	.007	−.034	−.018	.039	−.017	.038
DL-1M	−.204	−.134	−.198	−.051	−.200	−.065	−.048	−.039	−.051	−.005	−.051	−.008
DL-2M	−.217	−.131	−.204	−.065	−.207	−.080	−.030	−.045	−.045	.018	−.044	.015
DSN-M	−.221	−.126	−.203	−.073	−.206	−.088	−.024	−.053	−.046	.030	−.045	.027
EL	−.209	−.137	−.203	−.052	−.205	−.067	−.048	−.040	−.052	−.004	−.052	−.008
ID	−.209	−.145	−.208	−.045	−.210	−.060	−.038	−.020	−.034	−.015	−.035	−.017
IH	−.205	−.139	−.202	−.047	−.204	−.062	−.062	−.041	−.060	−.015	−.061	−.019
IL-1F	−.212	−.123	−.196	−.069	−.199	−.083	−.039	−.060	−.059	.024	−.058	.020
IL-1I	−.223	−.112	−.195	−.089	−.198	−.103	−.021	−.077	−.060	.057	−.058	.053
IL-1M	−.235	−.125	−.211	−.087	−.214	−.102	−.005	−.059	−.040	.053	−.038	.050
IL-2M	−.197	−.109	−.179	−.069	−.181	−.082	−.063	−.084	−.088	.026	−.087	.020
IL-2MR	−.227	−.128	−.207	−.077	−.210	−.092	−.022	−.056	−.047	.034	−.046	.031
N1R	−.240	−.144	−.225	−.072	−.228	−.089	.007	−.024	−.011	.030	−.010	.029
NATL-24/35	−.203	−.128	−.194	−.055	−.196	−.069	−.062	−.056	−.070	−.000	−.070	−.005
NATL25	−.194	−.099	−.171	−.076	−.174	−.089	−.074	−.103	−.106	.035	−.105	.027
RF	−.216	−.132	−.204	−.063	−.207	−.078	−.046	−.053	−.059	.010	−.059	.006
RH	−.214	−.112	−.191	−.081	−.194	−.095	−.038	−.080	−.071	.043	−.070	.038
SN-F	−.207	−.119	−.190	−.068	−.193	−.082	−.032	−.058	−.054	.028	−.053	.024
SN-I	−.223	−.146	−.217	−.056	−.219	−.072	−.021	−.022	−.025	.003	−.025	.001
SN-M	−.226	−.160	−.227	−.046	−.229	−.062	−.012	.003	−.005	−.013	−.005	−.014
SUB3	−.205	−.127	−.195	−.059	−.197	−.073	−.056	−.057	−.067	.006	−.067	.002
T-All	−.167	−.111	−.163	−.040	−.165	−.052	−.119	−.084	−.119	−.025	−.120	−.033
TH	−.201	−.140	−.200	−.042	−.202	−.057	−.056	−.032	−.051	−.019	−.052	−.023
TL12	−.222	−.137	−.211	−.063	−.213	−.078	−.014	−.030	−.027	.016	−.026	.014
TL-M	−.199	−.146	−.203	−.035	−.205	−.049	−.055	−.018	−.042	−.031	−.043	−.034
TL56	−.207	−.122	−.192	−.065	−.195	−.079	−.039	−.057	−.058	.020	−.057	.016
Overall	−.234	−.143	−.221	−.069	−.224	−.085	−.052	−.067	−.071	.019	−.071	.014

Note. Critical values for correlations are ±.065 at 5%, ±.085 at 1%, ±.092 at 0.5%, and ±108 at 0.1%.

Table 7 Correlation matrix of lexical target properties

	Priming	CELEX	HAL	SUB-WF	SUB-CD	Orth *N*	Phon *N*
Priming	1.000	−.109	−.091	−.155	−.137	−.151	−.163
CELEX	−.109	1.000	.772	.780	.800	.048	.108
HAL	−.091	.772	1.000	.780	.778	−.019	.073
SUB-WF	−.155	.780	.780	1.000	.985	.062	.166
SUB-CD	−.137	.800	.778	.985	1.000	.067	.149
Orth N	−.151	.048	−.019	.062	.067	1.000	.362
Phon N	−.163	.108	.073	.166	.149	.362	1.000
PhGr N	−.106	.018	−.041	.019	.022	.826	.364
OLD20	.215	−.168	−.074	−.188	−.196	−.637	−.430
PLD20	.242	−.177	−.096	−.217	−.215	−.385	−.672
BG Freq	−.054	.104	.117	.123	.132	.329	.019
Len Phon	.126	−.147	−.062	−.220	−.212	−.158	−.611
Len Syll	.135	−.083	.001	−.081	−.090	−.126	−.415
Homophones	−.074	.042	.011	.028	.013	.054	.433

	PhGr *N*	OLD20	PLD20	BG Freq	Len Phon	Len Syll	Homophones
Priming	−.106	.215	.242	−.054	.126	.135	−.074
CELEX	.018	−.168	−.177	.104	−.147	−.083	.042
HAL	−.041	−.074	−.096	.117	−.062	.001	.011
SUB-WF	.019	−.188	−.217	.123	−.220	−.081	.028
SUB-CD	.022	−.196	−.215	.132	−.212	−.090	.013
Orth N	.826	−.637	−.385	.329	−.158	−.126	.054
Phon N	.364	−.430	−.672	.019	−.611	−.415	.433
PhGr N	1.000	−.537	−.349	.201	−.139	−.169	.000
OLD20	−.537	1.000	.611	−.407	.380	.401	−.133
PLD20	−.349	.611	1.000	−.193	.665	.537	−.255
BG Freq	.201	−.407	−.193	1.000	.053	−.012	−.093
Len Phon	−.139	.380	.665	.053	1.000	.506	−.244
Len Syll	−.169	.401	.537	−.012	.506	1.000	−.183
Homophones	.000	−.133	−.255	−.093	−.244	−.183	1.000

Note. Critical values for correlations are ±080 at 10%, ±.096 at 5%, ±.125 at 1%, ±137 at 0.5%, and ±160 at 0.1%. CELEX = log. CELEX frequency (Baayen, Piepenbrock, & Gulikers, 1995); HAL = log. HAL frequency (Burgess, 1998); SUB-WF = log. subtitle frequency (Brysbaert & New, 2009); SUB-CD = log. subtitle contextual diversity (Adelman, Brown, & Quesada, 2006); Orth *N* = orthographic neighborhood size (Coltheart, Davelaar, Jonasson, & Besner, 1977); Phon *N* = phonological neighborhood size; PhGr *N* = phonographic neighborhood size (Adelman & Brown, 2007; Peereman & Content, 1997); BG Freq = mean bigram frequency; OLD20, PLD20 = orthographic/phonological Levenshtein distance, average of smallest 20 (Yarkoni, Balota, & Yap, 2008); Len Phon = length in phonemes; Len Syll = length in syllables; Homophones = total number of entries with same pronunciation.

two letters of seven-letter words (and no double substitution produced priming for five-letter words). It also contrasts with a variety of evidence from letter identification tasks showing the importance of exterior letters (in the form of higher accuracy of report; Estes, Allemeyer, & Reder, 1976). The condition that preserved only exterior letters, disrupting the identity of all four interior letters (EL: *dzbtkn*–DESIGN) produced no priming. Transpositions produced a different pattern, with initial adjacent transpositions (TL12: *edsign*–DESIGN) producing less priming than did final adjacent transpositions (TL56: *desing*–DESIGN); medial adjacent transpositions (TL-M: *desgin*–DESIGN) differed from neither. Against unrelated controls, Perea and Lupker also found that final and medial transpositions did not differ. Schoonbaert and Grainger, in contrast, found more priming for medial transpositions of five-letter words than for exterior transpositions. However, Schoonbaert and Grainger found no such difference with seven-letter words. Insertions and deletions produced a pattern such that disruptions of the final character (DL-1F and IL-1F: *desig*–DESIGN and *designl*–DESIGN) produced more priming than did medial (DL-1M and IL-1M: *dsign*–DESIGN and *desrign*–DESIGN) and initial (IL-1I: *mdesign*–DESIGN) disruptions. Given that insertions and deletions affect both letter identity and position, it would be consistent to suggest that the

Table 8 Regression coefficients in models predicting priming (average for all related prime types minus average for all control prime types) by target lexical properties

Neighbor measures:	All				Orth *N*, Phon *N*, PhGr *N*				Orth *N*, Phon *N*				OLD20, PLD20			
Frequency measure:	CELEX	HAL	SUB-WF	SUB-CD	CELEX	HAL	SUB-WF	SUB-CD	CELEX	HAL	SUB-WF	SUB-CD	CELEX	HAL	SUB-WF	SUB-CD
Intercept	−5.84	−0.65	4.98	2.55	20.22†	26.63*	31.12*	29.22 *	−1.43†				−1.10			
CELEX	−1.11				−1.41†					−1.36†				−1.12		
HAL		−1.14				−1.33†					−5.10 **				−4.41*	
SUB-WF			−4.40*				−5.01**					−4.76*				−3.96†
SUB-CD				−3.94†				−4.66*	0.43	0.67	−0.13	0.03	−2.53	−2.46	−3.07	−2.90
Len Phon	−2.71	−2.57	−3.14	−3.01	0.18	0.42	−0.34	−0.19	3.77	4.03	4.14	3.97	0.27	0.40	0.63	0.46
Len Syll	0.82	0.98	1.19	1.01	4.08	4.33	4.41	4.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BG Freq	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−1.54	−1.73	−1.97	−1.99	−1.21	−1.22	−1.45	−1.52
Homophones	−0.58	−0.73	−1.02	−1.00	−0.88	−1.07	−1.36	−1.37	−1.66†	−1.77 *	−1.71*	−1.68*				
Orth N	−1.61	−1.66	−1.61	−1.59	−2.70†	−2.80*	−2.65†	−2.64†	−0.43	−0.38	−0.37	−0.40				
Phon N	−0.06	−0.01	−0.01	−0.03	−0.50	−0.46	−0.44	−0.47								
PhGr N	1.95	1.93	1.80	1.84	1.78	1.76	1.62	1.65					8.44†	8.89†	8.28†	8.31†
OLD20	7.20	7.48	6.78	6.93									12.06**	12.01**	11.83**	11.96**
PLD20	11.94*	12.01*	11.82*	11.89*												
R-squared	7.64%	7.72%	8.40%	8.00%	5.25%	5.23%	6.10%	5.66%	5.04%	5.04%	5.93%	5.49%	7.34%	7.41%	8.12%	7.72%

pattern for these alterations is the combination of the patterns for the letter identity (substitution) and letter position (transposition) cases.

Insertions, deletions, and prime length

Putting aside the issue of position, even stimuli involving several deletions (SUB and DL-2M: *des*–DESIGN and *degn*–DESIGN) and insertions (IL-2M and IL-2MR: *desaxign*–DESIGN and *deshhign*–DESIGN) provided moderate priming relative to the unrelated baselines, and deletions (DL-1M: *dsgin*–DESIGN) were less disruptive than substitutions (SN-M: *desihn*–DESIGN). This is consistent with the patterns observed by Van Assche and Grainger (2006) and Norris, Kinoshita, and van Casteren (2010), respectively, despite their different prime and target lengths. Nevertheless, it is unclear to what extent the length of the unrelated baseline contributes to these results, since the length of the unrelated baseline (six letters) was not the same as that of the deletion and insertion primes. Thus, it will prove difficult to unambiguously disentangle the roles of stimulus length and insertion and deletion with the present data alone. Further constraint on theoretical accounts of these effects would come from studies systematically manipulating prime and target length.

Transposition versus replacement

Transposition primes produced more priming than did equivalent substitution primes, consistent with earlier findings (TL-M vs. DSN-M: *deisgn*–DESIGN vs. *dewvgn*–DESIGN; see, e.g., Perea & Lupker, 2003; Schoonbaert & Grainger, 2004). Indeed, a transposition involving a substituted letter (N1R: *dslign*–DESIGN) did not produce significantly less priming than did the substitution alone (SN-M: *desihn*–DESIGN), contrary to an earlier report with shorter (five-letter) stimuli (Davis & Bowers, 2006).

Extreme transpositions

Several of the extreme transpositions (T-All, RH, IH, TH: *edisng*–DESIGN, *sedngi*–DESIGN, *idgens*–DESIGN, *igndes*–DESIGN) produced small but significant priming effects. Previous studies had reported null effects from these types of primes (e.g., Guerrera & Forster, 2008; Lupker & Davis, 2009). The data are consistent with two possible (and not necessarily competing) explanations of the inconsistency. First, given the size of the effect, the previous null effects could arise from a lack of power. Second, the difference between studies could be due to the greater length of the stimuli in the earlier experiments, which means that the total amount of change from target to prime was greater in the prior studies (e.g., T-All was four transpositions in the prior studies but three transpositions in this one).

Unrelated baselines

The present study also used two different forms of unrelated baselines that have been employed inconsistently in the literature, one made up of arbitrary unrelated letters (*cbhaux*–DESIGN), and the other designed to form a pronounceable pseudoword (*voctal*–DESIGN). The latter led to faster responses and, hence, lower priming estimates. Faster responses for pseudoword unrelated primes (and hence, lower estimates) are what would be expected if the prime directly contributes to the word–nonword decision, because the pseudowords are more wordlike and, therefore, more suggestive of a "word" response. As such, although this result is not necessarily surprising,[5] it does point to yet another concern when comparing or agglomerating different experiments.

Uses of the database

The kinds of comparisons of conditions discussed above are not the primary basis for use of this database for establishing empirical patterns—since these comparisons are already all listed in Table 4—and, indeed, selecting subsets of the data to perform simple comparisons has not been how other mega-studies have been used (see Balota et al., 2012, for a review). Rather, mega-studies have been used to assess and compare models (e.g., Adelman & Brown, 2008b; Spieler & Balota, 1997), to consider the role of individual differences (Yap et al., 2012), and to investigate new (continuous) predictors or measures (e.g., Adelman & Brown, 2007; Adelman, Brown, & Quesada, 2006; Yarkoni et al., 2008). We discuss our preliminary findings along these lines below, and we envisage that many uses of these new data will be analogous to those with other mega-studies, with the expectation that predictors of interest will interact with type of prime.

Individual differences

The overall patterns of faster responses and less priming for those with better (written) language skills are consistent with the prior report of Andrews (2008). While good spelling was more strongly related to faster responding than was good vocabulary, there was no evidence that good spelling was associated with less facilitatory priming over and above the effect of language competence in general, the pattern reported by Andrews and Lo (2012).

[5] On the other hand, there are reasons to suppose the wordlikeness of the prime might not contribute to the word–nonword decision directly. First, the pseudoword primes were just as nonword-like as the foils. Second, the foils were very wordlike, disfavoring a criterion based on wordlikeness rather than identification of a single word.

Target differences

A moderator of whether a target could be primed by an orthographically similar prime that was identified in one of the earliest studies in the paradigm (Forster, Davis, Schoknecht, & Carter, 1987) is the neighborhood size of the target. Here, we found that of the neighborhood variables that predicted priming, the strongest, and the only one accounting for unique variance, was PLD20. This phonological measure stands in contrast to the orthographic neighborhood variables that are normally of interest in the context of this paradigm. One possible interpretation is that lexical decisions are made at the phonological level (e.g., Rastle & Brysbaert, 2006), and so even orthographic priming is sensitive to phonological competition. Alternative interpretations could suggest that PLD20 is a better indicator of the truly relevant orthographic neighborhood because that orthographic neighborhood is sensitive to multiletter graphemes, rather than letters, or that the consonant–vowel pattern is important.

Whether the amount of priming obtained is sensitive to the frequency of the target has been a subject of some debate, with many experiments (e.g., Forster & Davis, 1984) finding no such effect but others, indeed, finding the effect when the manipulation is sufficiently large and the word stimuli are all familiar to participants (e.g., Kinoshita, 2006). Here, only with one frequency count—albeit the one that is most predictive of lexical decision times (Brysbaert & New, 2009)—did we reliably find an effect on priming such that higher frequency words were less susceptible to priming. Even if this means that the effect is a real one, it also means that it is quite a weak effect (possibly, in part, because of restricted range). Given that the effect of frequency is so weak and frequency and neighborhood variables are correlated, it would not be so surprising if there were some other yet-to-be-constructed neighborhood variable that would subsume the effect of both PLD20 and frequency.

Conclusion

The present database is the first of its kind for investigating orthographic (masked form) priming. It should serve as a benchmark data set in a variety of investigations surrounding orthographic processing. An important example is the analysis of differences in mean priming for different prime types and the implications for models. Other uses to which these data could be put include moderation by individual differences, moderation by properties of items, and sequential effects and variability.

Acknowledgments James S. Adelman and Colin J. Davis are the organizers (senior authors) of this project. Other authors are listed in order of contribution. We thank Alexa Banculli, Peter Bowles, Josef

Broder, Helen Brown, Leo Chong, Alison Crumpton, Laura Cunniffe, Louise Goddard, Laura Grima, Lydia Gunning, Anna Hall, Hyun Kyoung Jung, Hannah Jenkins, Julie Lee, Bree Loethen, Mayumi Kohiyama, Stuart Miller, Luke Mills, Nicole Newson, Ann-Marie Raphail, Emma Roscow, Jocelyn Schock, Annabel Snell, Allison Teevan, Wan Zhen Chua, Louise Warner, Melissa Yeo, and Jimmie Zhang for their assistance in data collection and Dave Balota for helpful comments. Data collection at Royal Holloway, University of London, was supported by a grant from the Experimental Psychology Society.

References

Adelman, J. S. (2011). Letters and time in retinotopic space. *Psychological Review, 118*, 570–582.

Adelman, J. S., & Brown, G. D. A. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review, 14*, 455–459.

Adelman, J. S., & Brown, G. D. A. (2008a). Methods of testing and diagnosing models: Single and dual route cascaded models of word naming. *Journal of Memory and Language, 59*, 524–544.

Adelman, J. S., & Brown, G. D. A. (2008b). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review, 114*, 214–227.

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*, 814–823.

Andrews, S. (2008). Lexical expertise and reading skill. *Psychology of Learning and Motivation, 49*, 249–281.

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General, 139*, 299–318.

Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 152–163.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (Release 2) [CD-ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459.

Balota, D. A., Yap, M. J., Hutchison, K. A., & Cortese, M. J. (2012). Megastudies: What do millions (or so) of trials tell us about lexical processing? In J. S. Adelman (Ed.), *Visual word recognition* (Models and methods, orthography and phonology, Vol. 1, pp. 90–115). Hove: Psychology Press.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.

Burgess, C. (1998). From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavior Research Methods, Instruments and Computers, 30*, 188–198.

Burt, J. S., & Tate, H. (2002). Does a reading lexicon provide orthographic representations for spelling? *Journal of Memory and Language, 46*, 518–543.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornič (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Damian, M. F. (2010). Does variability in human performance outweigh imprecision in response devices such as computer keyboards? *Behavior Research Methods, 42*, 205–211.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117*, 713–758.

Davis, C. J., & Bowers, J. S. (2006). Contrasting five different theories of letter position coding: Evidence from orthographic similarity effects. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 535–557.

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

Estes, W. K., Allemeyer, D. H., & Reder, S. M. (1976). Serial position functions for letter identification at brief and extended exposure durations. *Perception & Psychophysics, 19*, 1–15.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680–698.

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

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers, 35*, 116–124.

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

Grainger, J., & van Heuven, W. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The mental lexicon* (pp. pp. 1–pp. 24). New York: Nova Science.

Guerrera, C., & Forster, K. (2008). Masked form priming with extreme transposition. *Language and Cognitive Processes, 23*, 117–142.

Humphreys, G. W., Evett, L. J., & Quinlan, P. T. (1990). Orthographic processing in visual word identification. *Cognitive Psychology, 22*, 517–560.

Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2007). Predicting semantic priming at the item level. *Quarterly Journal of Experimental Psychology, 61*, 1036–1066.

Hutchison K. A, Balota D. A, Neely J. H, Cortese M. J, Cohen-Shikora E. R, Tse C. -S, Buchanan E, (2014). The semantic priming project. Behavior Research Methods. doi: 10.3758/s13428-012-0304-z

Kinoshita, S. (2006). Additive and interactive effects of word frequency and masked repetition in the lexical decision task. *Psychonomic Bulletin & Review, 13*, 668–673.

Lupker, S. J., & Davis, C. J. (2009). Sandwich priming: A method for overcoming the limitations of masked priming by reducing lexical competitor effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 618–639.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375–407.

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

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.

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*, 528–530.

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language, 37*, 382–410.

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.

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology, 53*, 97–145.

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.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Journal of Psychology, 9*, 371–377. doi:10.1080/00223980.1940.9917704

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8*, 411–416.

Ulrich, R., & Giray, M. (1989). Time resolution of clocks: Effects on reaction time measurement—good news for bad clocks. *British Journal of Mathematical and Statistical Psychology, 42*, 1–12.

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.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 53–79.

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