



The influence of place and time on lexical behavior: A distributional analysis

Brendan T. Johns¹ · Randall K. Jamieson²

Published online: 26 August 2019
© The Psychonomic Society, Inc. 2019

Abstract

We measured and documented the influence of corpus effects on lexical behavior. Specifically, we used a corpus of over 26,000 fiction books to show that computational models of language trained on samples of language (i.e., subcorpora) representative of the language located in a particular place and time can track differences in people’s experimental language behavior. This conclusion was true across multiple tasks (lexical decision, category production, and word familiarity) and provided insight into the influence that language experience imposes on language processing and organization. We used the assembled corpus and methods to validate a new machine-learning approach for optimizing language models, entitled *experiential optimization* (Johns, Jones, & Mewhort in *Psychonomic Bulletin & Review*, 26, 103–126, 2019).

Keywords Lexical organization · Lexical semantics · Distributional semantics · Big data · Machine learning

Bartlett (1928, 1932) conducted a classic experiment in cognitive psychology to examine how people remember. In the experiment, participants read an Indigenous American story entitled “The War of the Ghosts.” When prompted to recall the story, people inserted their own knowledge. For example, participants used the word *boat* (i.e., a word that they had experience with) in place of *canoe* (i.e., a word they likely did not know). Although Bartlett’s demonstration is remembered as a foundational example in the theory of reconstructive memory, it also makes another point.

When people read or hear language, they comprehend that language through the lens of their own experience. For example, when asked to play a game of football, a person’s interpretation of that request might change depending on the side of the Atlantic where the person was raised. Similarly, if one is asked to play a game of “roque,” only people familiar with croquet variants from the late 1800s would understand the rules.

But do the subtler differences in language experience exert a meaningful and distinguishable influence on people’s behavior and cognition?

By traditional methods of scholarship, a complete examination of language experience is intractable. However, recent advances in natural language processing, coupled with the availability of sizeable text corpora, have changed the game (e.g., Brysbaert, Mandera, & Keuleers, 2018; Chubala, Johns, Jamieson, & Mewhort, 2016; Green, Feinerer, & Burman, 2013, 2015; Johns, 2019; Johns, Jones, & Mewhort, 2019; Johns, Mewhort, & Jones, 2019; Johns & Jones, 2015; Jones, 2017; Jones, Dye, & Johns, 2017; Landauer & Dumais, 1997). For example, Johns and Jamieson (2018) applied theories of natural language processing to analyzing language use in published fiction. On the basis of the analysis, they reported that language differs in meaningful and measurable ways between genres, books, and authors. But, from a theoretical perspective, their results set up a more interesting problem: Can theories of natural language processing track differences in language behavior conditional on language experience?

The question has been analyzed previously in several fields, including sociolinguistics, corpus linguistics, and psycholinguistics (e.g., Baker, 2010; Biber, 1993; Brysbaert, Keuleers, & New, 2011). For example, Brysbaert and New (2009; SUBTLEX) recorded modern word frequency norms based on language from American film and television, whereas van Heuven, Mandera, Keuleers, and Brysbaert (2014; SUBTLEX-UK) recorded modern word frequency norms based on language in British film and television. Although the details vary, the authors of those projects arrived at a general conclusion germane to the present investigation: Norms

✉ Brendan T. Johns
btjohns@buffalo.edu

¹ University at Buffalo, Buffalo, NY, USA

² University of Manitoba, Winnipeg, Manitoba, Canada

developed from local sources (i.e., the same country) do a better job of tracking people's language behavior than do norms developed from a different country. Given the success of the SUBTLEX norms, this approach to building word frequency datasets has been extended to a number of other languages, including Chinese (Cai & Brysbaert, 2010), Dutch (Keuleers, Brysbaert, & New, 2010), and Polish (Mandera, Keuleers, Wodniecka, & Brysbaert, 2015).

Our goal was to expand on this line of investigation. Our first goal was to use fiction books written by authors of known nationality and era to measure and characterize differences in language use as a function of geography and time (although, as Brysbaert & New, 2009, point out, subtitles may provide a more naturalistic type of language; see Herdağdelen, & Marelli, 2017, for similar arguments about using word frequency values from social media sources). Our second goal was to determine whether lexical models trained on those place- and time-specific subcorpora better explain people's lexical behavior when the nationality and era of the subcorpus matches the nationality of the participants and the era in which the data were collected—a proposition we call the *selective-reading hypothesis*. Our third goal was to use the assembled subcorpora and behavioral norms to evaluate the technique of experiential optimization (EO; Johns, Jones, & Mewhort, 2019), a new machine-learning method that uses lexical data to infer and identify sections of the corpus that match the experimental participants' language experience.

Corpus development and analysis

In previous work, we reported that language varies in meaningful ways conditional on authors and genres (Johns & Jamieson, 2018). However, we did not study differences as a function of when authors were born or the places in which they lived. Given that our goal was to show that language behavior reflects the language environment, we first repeated and confirmed that analysis using a larger corpus of books written by American and British authors in different years.

To support the analysis, we assembled a corpus of 26,000 books from the internet. We then used the corresponding metadata to tag each book by its author, the author's place of birth, the author's date of birth, and the book's genre. We also collected information from those sources on book length (i.e., number of words). All information was obtained from the websites Goodreads, Amazon, and Wikipedia.

Table 1 provides information about the books in our corpus, which in total includes over two billion words. As is shown there, the total corpus includes over 26,000 books written by over 3,000 different authors: over 1.3 billion words written by 1,999 American authors, and about 500 million words written by 738 British authors. More specific information is included in Table 1, which differentiates books by

author place of birth and genre, where author genre was determined by the most frequent genre in which an author published.¹

We also recorded dates of birth for 2,088 of the 3,209 authors in the corpus, yielding a range of birth dates between 1801 and 1998 and an overall corpus size of approximately 1.5 billion words for the time-delineated corpus. Figure 1 shows the numbers of words in the sample as a function of author birth date. As the figure shows, most words in the sample were written by authors who were born between 1925 and 1975. Therefore, we split the sample so that authors born before 1942 were included in an *old-generation* subcorpus of 762 million words, whereas authors who were born from 1942 onward constituted a *new-generation* subcorpus of 792 million words. Those subcorpora will be used to characterize language use as a function of time (i.e., old vs. new).²

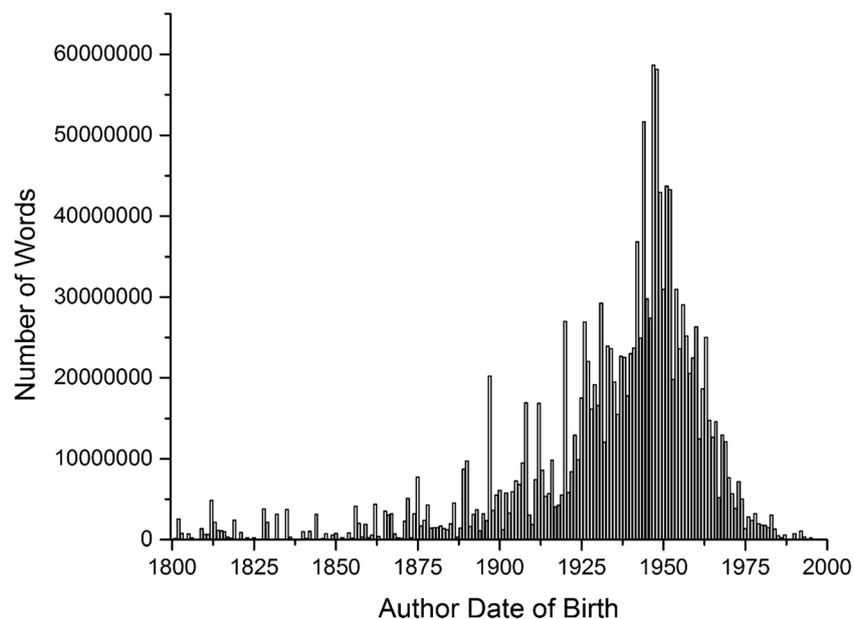
For the first analysis, we used Johns and Jamieson's (2018) methods to measure the similarity (and thus the differences) in language use between books written by authors in the United States (USA) and United Kingdom (UK) subcorpora and books written by authors in the old-generation versus new-generation subcorpora. The method involved, first, identifying the 100,000 highest-frequency words in both of the relevant subcorpora (i.e., USA and UK subcorpora for the place analysis, and old- and new-generation subcorpora for the time analysis). Second, we constructed an Author (i.e., 3,209 authors in the place comparison and 2,088 in the time comparison) × Word (100,000 highest-frequency words) matrix that recorded the number of times that each of the 100,000 highest-frequency words appeared in each author's books. Third, we converted all word frequencies to their log equivalents (i.e., $n' = \ln[n + 1]$), where n is the word count from the book). Finally, we computed the cosine similarity of each author's vocabulary vector to every other author's vocabulary vector. If two authors had a date of birth within a 30-year window (a common method of calculating generational differences; see Tremblay & Vézina, 2000), they were classified as belonging to the

¹ Although this is less precise than the methods used in other studies examining the impact of genre on writing (see Johns & Jamieson, 2018), tagging each book by its genre was not practically feasible for such a large collection.

² As Fig. 1 shows, the distribution of the date-of-birth information is roughly Gaussian, centered around the mid-1940s, with a negative skew. There are a number of reasons for this distribution. The first reason is that a book had to be available as an e-book before being translated into a machine-readable format. For older books, this limited the sample to classics, or at least to books of interest to the current generation. Second, in order to attain date-of-birth information for an author, there needed to be publicly available biographical information for that author (this also applied to the availability of place-of-birth information). This again constrained our sample and led to fewer authors being included in our sample; even if biographical information was available for contemporary authors, their date-of-birth information might not be determinable.

Table 1 Characteristics of the book collections for authors born in American and Britain

Author's country of birth	Genre	Number of authors	Number of books	Words per book	Total number of words
America	Fantasy	185	1,492	97,959	146,154,828
	Historical fiction	100	627	93,242	58,462,734
	Mystery	213	1,628	72,716	118,381,648
	Romance	458	3,729	78,591	293,065,839
	Thriller	108	813	107,270	87,210,510
	Young adult	178	1,707	39,278	67,047,546
	Crime	37	229	78,525	17,982,225
	Literature	116	938	94,994	89,104,372
	Horror	51	558	82,874	46,243,692
	Science fiction	332	4,088	74,368	304,016,384
	Christian	44	366	66,427	24,312,282
	Western	41	407	63,409	25,807,463
	Nonfiction	136	533	101,022	53,844,726
	<i>Total/Average</i>	1,999	17,115	80,821	1,383,254,048
Britain	Fantasy	41	497	97,311	48,363,567
	Historical fiction	79	579	118,323	68,509,017
	Mystery	74	679	74,584	50,642,536
	Romance	84	697	66,463	46,324,711
	Thriller	43	326	86,345	28,148,470
	Young adult	81	613	42,805	26,239,465
	Crime	34	353	68,987	24,352,411
	Literature	148	1,184	86,926	102,920,384
	Horror	17	140	101,366	14,191,240
	Science fiction	70	765	80,408	61,512,120
	Christian	1	8	63,337	506,696
	Western	4	80	45,529	3,642,320
	Nonfiction	62	250	119,520	29,880,000
	<i>Total/Average</i>	738	6,171	80,916	499,330,737
Other nationalities	<i>Total/Average</i>	472	3,352	76,324	257,275,867

**Fig. 1** Numbers of words in the book sample, organized by author date of birth

same generation; otherwise, they were classified as belonging to different generations.

The middle panel in Fig. 2 shows vocabulary similarity between authors with the same versus different nationalities (i.e., both USA or both UK vs. one USA and one UK). As is shown, the vocabularies of authors born in the same country were more similar than the vocabularies of authors born in different countries.

The bottom panel in Fig. 2 shows vocabulary similarity between authors from the same versus different generations (i.e., less than vs. more than 30 years' difference). As the figure shows, the vocabularies of authors belonging to the same generation were more similar to one another than the vocabularies of authors belonging to different generations.

To provide an intuitive point for comparison, the top panel of Fig. 2 shows vocabulary similarity for authors writing in the same versus different genres, independent of their nationalities and ages. As is shown, the

vocabularies of authors writing in the same genre were more similar to one another than the vocabularies of authors writing in different genres, an expected result (Johns & Jamieson, 2018). More importantly, the scale of that difference provides an intuitive picture for assessing the difference in vocabulary similarity distributions as a function of place and time.

Naturally there are trends in books, so there is a chance that differences in vocabulary use conditional on place and time might be conflated with those trends (e.g., the popular genres at a given time). Therefore, we recomputed the time and place analyses for authors depending on whether they wrote in the same versus different genres. As is shown in Fig. 3, the pattern of similarities in Fig. 2 was preserved. Authors of the same nationality and authors belonging to the same generation used more similar vocabularies, even when they were writing in different genres.

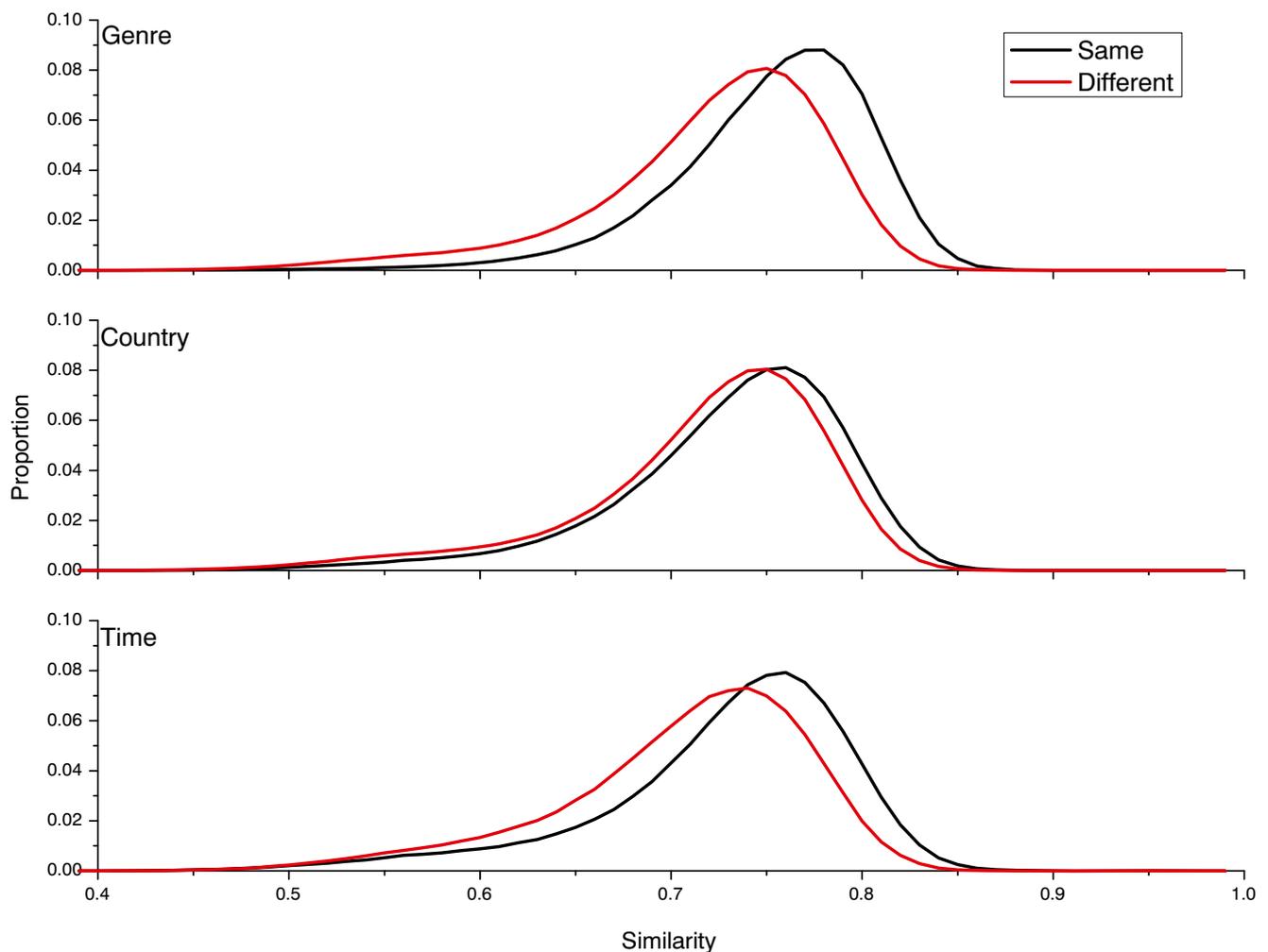


Fig. 2 Histograms of similarity distributions for authors who wrote in the same or different genres (top panel), had the same or different places of birth (middle panel), and who were born in the same or in different

generations (bottom panel). The genre comparison serves as an intuitive comparison for the differences seen in the bottom two panels

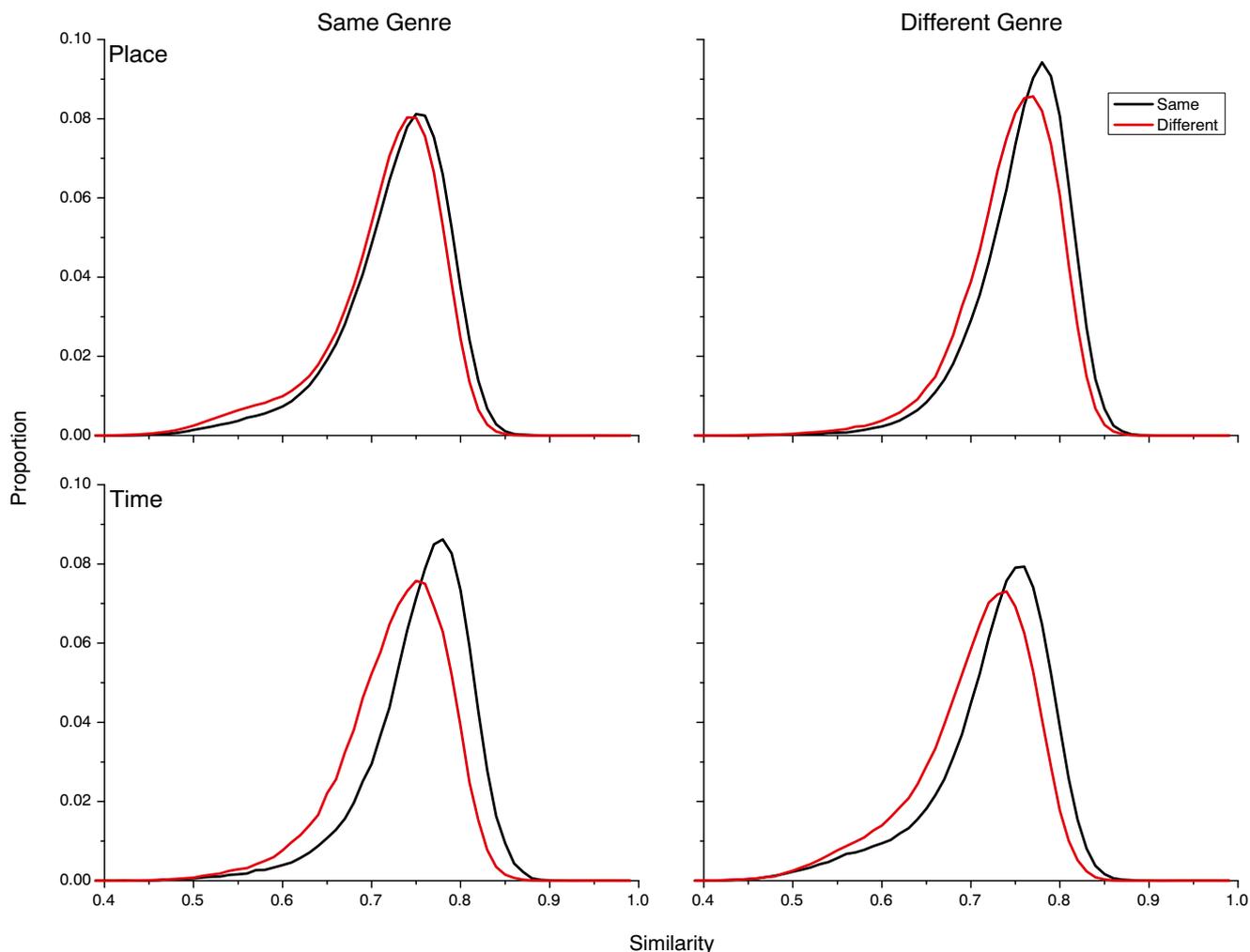


Fig. 3 Histograms of authors who wrote in the same or different genres and had the same or different places of birth (top) or who were born in the same or different generations (bottom)

On the basis of the analysis, we concluded that vocabulary varies systematically as a function of both time and place, even when authors write in different genres. The results corroborate Johns and Jamieson's (2018) results, using a different and substantially larger corpus.

Method

Although we showed differences in vocabulary as a function of author nationality and date of birth, it was an open question whether those differences would be reflected in the lexical behavior of experimental participants. To answer the question, we now examined the correspondence (or lack thereof) between variation in written language in our subcorpora as a function of time and place and the lexical behavior of participants collected in the USA and the UK in different years.

Empirical databases

Both lexical organization and lexical semantic data were tested. To examine lexical semantics, semantic category production data were used. In a semantic category production task, participants were cued with a category label (e.g., *vegetable*) and had to produce as many examples from that category as possible (e.g., *carrot, lettuce, cucumber*). By tradition, the frequency of production is interpreted as a snapshot of the psychological structure for the mental category (e.g., Battig & Montague, 1969; Rosch & Mervis, 1975). Three datasets were examined: the classic Battig and Montague norms, the updated norms of Van Overschelde, Rawson, and Dunlosky (2004), and the British norms of Hampton and Gardiner (1983). The main time comparison was between the Battig and Montague and van Overschelde et al. norms. The main place comparison was between all three datasets, since both the Battig and Montague and Van Overschelde norms were

collected in the USA, whereas the Hampton and Gardner data were collected in the UK.

The words included in the analysis were reduced so as to only include words that appeared across datasets, to ensure that the difference in production values, and not the different words in the norms, were causing the different fits to the data. For the time comparison between the Battig and the Van Overschelde norms, this reduction resulted in 40 categories and 803 exemplars. The same reduction was done to compare the Hampton norms to the Battig and Van Overschelde norms, which resulted in 178 words across 11 categories for the place comparison.

To explore lexical organization, word familiarity and lexical decision data were used. In a word familiarity task, participants are simply asked to rate how familiar they are with a given word on a scale. Three datasets had been collected from North America: the norms of Paivio, Yuille, and Madigan (1968; familiarity data later released by Paivio, 1974); the norms of Stratton, Jacobus, and Brinley (1975; $n = 543$); and the extended Paivio et al. (1968) norms of Clark and Paivio (2004). The Paivio (1974) and Clark and Paivio norms were collected in Canada, but given the cultural overlap between Canada and the USA, these three datasets could be considered to be representative of the USA. Paivio et al. contained 925 words, and Clark and Paivio collected data on these same words, so these words were used to directly compare the datasets. These were contrasted with two datasets collected in the UK: the norms of Gilhooly and Logie (1980, $n = 1,944$) and the norms of Stadthagen-Gonzalez and Davis (2006, $n = 1,526$). Given the publication dates of these datasets and the locations where they were collected, these different datasets provided a powerful basis for testing our hypothesis.

Additionally, two mega-datasets of lexical decision data were used to examine the influence of place on lexical behavior: the English Lexicon Project of Balota et al. (2007), and the British Lexicon Project of Keuleers, Lacey, Rastle, and Brysbaert (2012). The data from the English Lexicon Project were collected from sites around the USA, whereas the British Lexicon Project data were collected from participants in Great Britain. The English Lexicon Project database contains 40,481 words, whereas that for the British Lexicon Project contains 28,515 words. To design a maximally inclusive but balanced comparison, we computed correlations over the 16,214 words that appeared in both the American and British lexical decision databases (a sample that also excluded any words that participants identified as words worse than chance: 50% accuracy). The data used were z -transformed lexical decision times.

Models

We used a simple word frequency model to explore the word familiarity and lexical decision data (for a recent review of word frequency, see Brysbaert et al., 2018). Although more complex models of word frequency have been developed (e.g., semantic diversity counts; Jones, Johns, & Recchia, 2012), using simple word frequency allowed for an uncomplicated examination of whether the word frequencies observed in the different subcorpora would map onto word familiarity ratings and lexical decision times as a function of time and place.

To account for the category production data, we used the BEAGLE model of semantics (Jones & Mewhort, 2007). Broadly, BEAGLE works by “reading” a text corpus and, en route, encoding each word’s meaning into a set of corresponding vectors. The theory is one in a larger class, labeled *distributional models of language* (e.g., Griffiths, Steyvers, & Tenenbaum, 2007; Landauer & Dumais, 1997).

To train BEAGLE, each of the i unique words in a corpus is represented by a unique n -dimensional environment vector, e_i , with each element being assigned a random deviate from a normal distribution with mean zero and variance $1/n$. In the simulations that follow, dimensionality was set to $n = 1,024$. Environment vectors are stable over a simulation and are meant to serve as unique identifiers for the words in the corpus.

Next, the model “reads” the corpus one sentence at a time, to build a semantic memory vector, m_i , for each word. The memory vector for each word is composed of two kinds of information: *context information* and *order information*.

Context information is computed by summing the environmental vectors for all other words in the same sentence (i.e., excluding the word of interest) into the representation for that word. The summing of environmental vectors in this manner causes the memory vectors for all words in the same sentence to grow more similar to one another.

Order information encodes how a word is used within a sentence and is computed by encoding all of the n -grams (up to a specified size) that a word is part of within a sentence.³ The computation of order information relies on noncommutative circular convolution (Plate, 1995) to bind the environmental vectors into unique n -gram vectors, which are then summed into the target word’s order representation. The order representation encodes how a word is used in relation to the words that surround it within a sentence.

In sum, a word’s context vector represents pure co-occurrence information, whereas order information encodes a simplified representation of syntactic relations. The representation used here was the sum of the context and order

³ The standard is to use n -grams up to size 7 (Jones & Mewhort, 2007), which were also used here.

vectors. Despite its simplicity, BEAGLE explains a broad range of semantic and language behavior (e.g., Jones & Mewhort, 2007; Recchia, Sahlgren, Kanerva, & Jones, 2015).

Results

Word familiarity

Our first analysis tested whether word frequencies tabulated from the new-books corpus (i.e., authors born after 1942) accounted for the word familiarity data in the database of newer experimental norms (i.e., Clark & Paivio, 2004; Stadthagen-Gonzalez & Davis, 2006) better than for the older experimental norms (i.e., Gilhooly & Logie, 1980; Paivio et al., 1968; Stratton et al., 1975), and vice versa. The top panel of Fig. 4 shows the correlations between the old-book and new-book subcorpora relative to the word familiarity norms in the old and new empirical sets of norms. As is shown, the word frequency values computed from the old- and new-book subcorpora provide better fits to their time-appropriate empirical norms: The old-book corpus matches word familiarity ratings better in the old sets of norms, and the new-book corpus matches word familiarity ratings better in the new sets of norms better.

Of course, there is a great deal of shared variance between word familiarity ratings in the old and new experimental norms, and that shared variance might work against identifying the differences between the two for predicting performance. Thus, to measure the predictive power of the differences in word familiarity in the old and new norms, we applied regression to quantify the unique variance accounted for in the word familiarity norms by word frequency in the old- and new-book subcorpora. The analysis is standard and provides a measure of the predictive gain (i.e., measured as the percent Δ of R^2 improvement) for one predictor over a competing predictor (see Adelman, Brown, & Quesada, 2006; Johns, Gruenenfelder, Pisoni, & Jones, 2012; Johns, Sheppard, Jones, & Taler, 2016; Jones et al., 2012).⁴

The bottom panel of Fig. 4 shows the results of the regression. As can be seen there, word frequencies tabulated from the old-book corpus (i.e., books written by authors born before 1942) account for nearly all the unique variance in the old empirical norms, whereas word frequencies tabulated from the new-book corpus (i.e., books written by authors born in

or after 1942) accounts for nearly all of the unique variance in the more recent empirical norms. These results provide a first, positive validation of the selective-reading hypothesis (and follow similar results found by Taler, Johns, & Jones, 2019, in a large-scale examination of verbal fluency performance).

A second test of the selective-reading hypothesis was to determine whether place also provides a unique signature in lexical behavior. The top panel of Fig. 5 shows correlations between word frequencies tabulated from our USA and UK subcorpora relative to word familiarity norms collected in North America (Clark & Paivio, 2004; Paivio et al., 1968; Stratton et al., 1975) and the UK (Gilhooly & Logie, 1980; Stadthagen-Gonzalez & Davis, 2006). The bottom panel shows regression results for the amount of unique variance accounted for by each corpus. As can be seen, the same pattern is observed as a function of place that was observed as a function of time: The USA book corpus matches word familiarity ratings better for the North American norms, and the UK book corpus matches word ratings better for the UK norms. This result corroborates our time-based analysis and provides additional support for the selective-reading hypothesis.

Lexical decision data

The correlations of word frequency values to the data from the English and the British Lexicon Projects are shown in the top panel of Fig. 6. The bottom panel shows the unique variance that the two frequency sets explain. This figure shows a result very similar to that from the word familiarity dataset. Word frequency in the USA books corpus accounted for more variance in lexical decision times collected in the USA than in the lexical decision times collected in the UK, and the opposite was true for the UK data.

The finding of the advantage of a corpus written by authors born in the UK replicates the finding of van Heuven, Mandera, Keuleers, and Brysbaert (2014), who found a similar increase in variance accounted for when using subtitle files from UK-based television and films to account for lexical decision data collected in the UK. However, the advantage for the corpus of books from the USA for the English Lexicon Project data is substantially larger than those from other studies attempting to model these data (e.g., Adelman et al., 2006; Jones et al., 2012). It is worth pointing out that this is not due to our book corpus providing an overwhelmingly better fit to those data than other subcorpora (the raw correlations are consistent with the other corpora used). Rather, our UK subcorpus provides a relatively poor fit to those data—a difference that enables a bigger advantage for the book corpus from authors born in the USA (this point also applies to the familiarity and category norm data). Finally, we speculate that the disparity in fits between the English and British Lexicon Projects in Fig. 6 might be due to an asymmetry in reading experience: American students likely have less experience with

⁴ Only word frequency values were included in the regression. The methodology has been used extensively to distinguish between different, highly correlated models of lexical organization (e.g., contextual diversity vs. word frequency; Adelman et al., 2006) and has been shown to be reliable for distinguishing the unique contributions of different metrics. Given that the word frequency measurements from the corpora used here are also highly correlated, we are confident that the results found here are consistent with those of past studies.

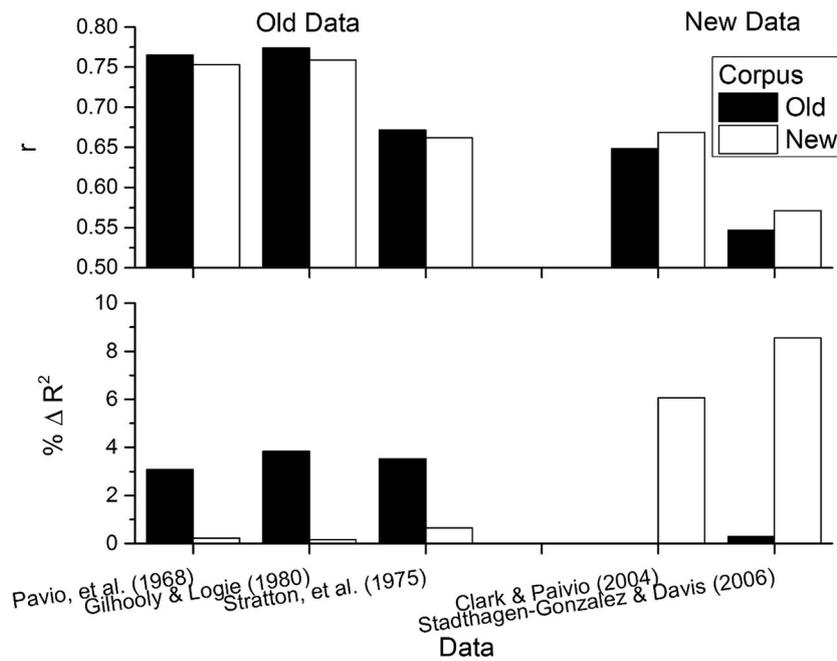


Fig. 4 The top panel displays the correlations of word frequency values derived from the old and new subcorpora to older versus more recent familiarity norms. The bottom panel displays the amounts of unique variance accounted for by the word frequency counts

British authors than British students have with American authors. However, this hypothesis needs to be tested in future research.

Category production norms

To generate predictions for category production from the BEAGLE model, we computed the cosine similarity of the

vector that corresponded to a given category label (e.g., *vegetable*) to the vectors that corresponded to each of the relevant category exemplars (e.g., *carrot*, *lettuce*, *cucumber*). Then we converted the cosines to ranks and computed the Spearman rank correlation between the exemplar ranks generated from the model and the corresponding exemplar ranks in the data (i.e., Battig & Montague, 1969; Rosch & Mervis, 1975).

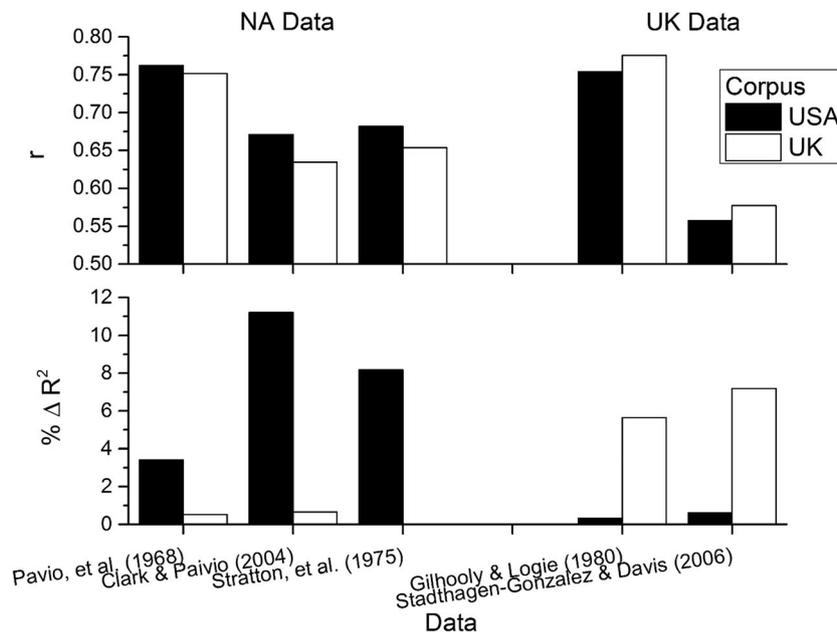


Fig. 5 The top panel displays the correlations of word frequency values derived from the USA and UK subcorpora to word familiarity norms collected in the USA versus the UK. The bottom panel displays the amounts of unique variance accounted for by the word frequency counts across the different sets of word familiarity norms

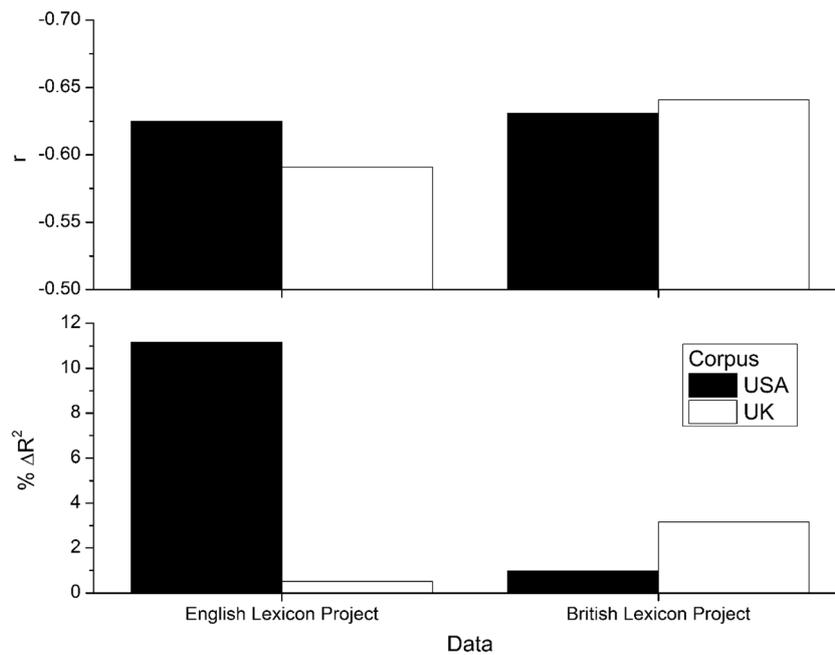


Fig. 6 The top panel displays the correlations of word frequency values derived from the USA and UK subcorpora to lexical decision data collected in the USA (English Lexicon Project) versus the UK (British

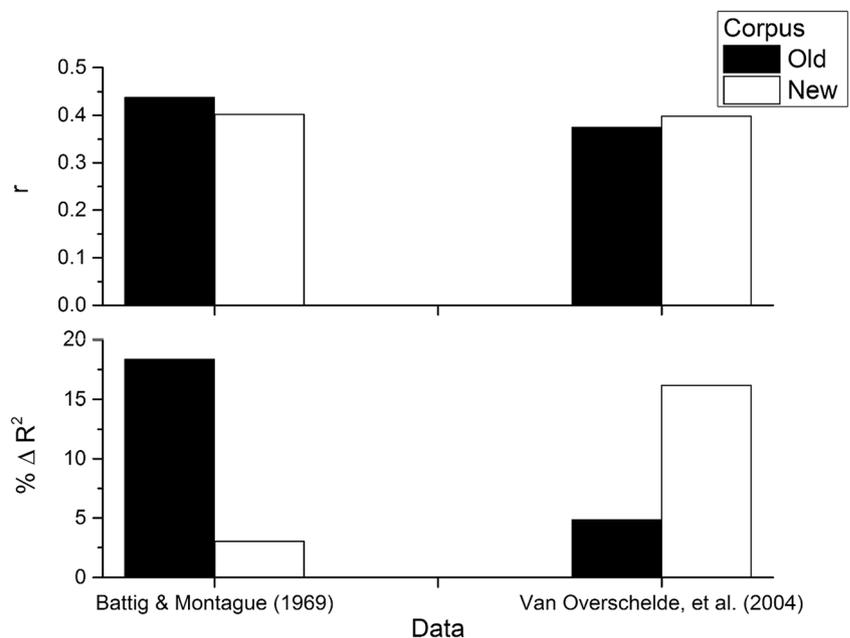
Lexicon Project). The bottom panel displays the amounts of unique variance accounted for by the two word frequency models across the different sets of norms

The first analysis focused on the effect of training the model on time-specific subcorpora. Figure 7 shows the correlations between the two models and category production frequency (top panel) and the amount of unique variance each model accounted for (bottom panel). As is shown there, the results for the category production data as a function of time mirror the results for the word familiarity data: Semantic

vectors derived from the old-book corpus explains the older Battig and Montague (1969) production norms better, whereas the semantic vectors derived from the new-book corpus explains the newer Van Overschelde et al. (2004) norms better.

The results from our final analysis, presented in Fig. 8, tested the influence of place on semantic category production. As expected, the results as a function of place mirrored the

Fig. 7 The top panel displays the fit of the BEAGLE model of semantics derived from the old and new subcorpora to older and more recent category production data. The bottom panel displays the amounts of unique variance accounted for by the two semantic models.



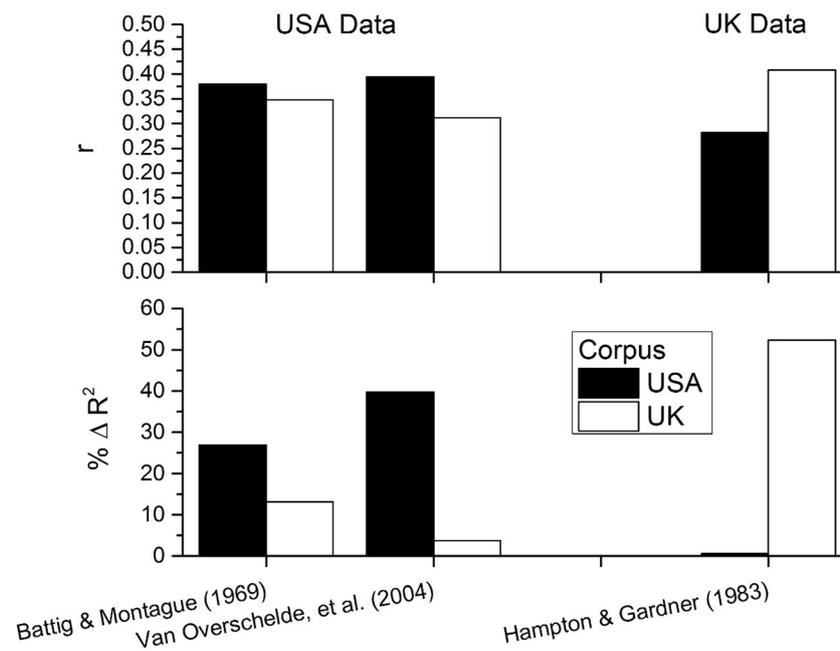


Fig. 8 The top panel displays the fit of the BEAGLE model of semantics trained on the USA and UK book subcorpora to category production norms collected in the USA versus the UK. The bottom panel displays

the amounts of unique variance accounted for by the two semantic models across the different sets of norms

results as a function of time in Fig. 7: Semantic vectors derived from the USA book corpus predicted the American Battig and Montague (1969) and Van Overschelde et al. (2004) production norms better, whereas the semantic vectors derived from the UK book corpus predict the British Hampton and Gardner (1983) norms better. The results provide additional support in favor of the selective-reading hypothesis: Training a semantic memory model on a country-appropriate corpus offers the best fit to category production norms collected in that country.

Discussion

Our goal was to assess the selective-reading hypothesis across several classes of lexical behavior. We found that the time- and place-matched subcorpora yielded a large and systematic advantage over the time- and place-unmatched subcorpora. The advantage was consistent for word familiarity, lexical decision, and category production data, and suggests that differential language experience exerts a measurable influence on language processing. It also demonstrates a subtler message: The language sample that a language model is trained on impacts its ability to account for lexical behavior. Specifically, if the materials that a corpus-based model is trained on mismatch the lexical experiences that a group of participants have had, that model might be rejected not because it is a poor model, but because it has a mismatched language background. In the next section of this article, we address the issue of matching language experience through a

new machine-learning method called experiential optimization.

A validation of experiential optimization

The analyses above show that the lexical behavior of participants across several behavioral tasks was sensitive to the time and place that the behavior was collected in. The combined results suggest that one way to construct better, more powerful models of lexical behavior is to train models with appropriate linguistic materials that accurately represent the language experience of the experimental participants.

Recently, Johns, Jones, and Mewhort (2019) presented a new machine-learning framework that is designed to accomplish just this. Their method, called experiential optimization (EO), uses the inherent variability in language (see Johns & Jamieson, 2018) to optimize models of natural language to sets of human behavior. Johns, Jones, and Mewhort showed that EO allows for benchmark fits across a diverse number of areas in the study of language and memory, including lexical organization, lexical semantics, sentence processing, and episodic recognition. In doing so, they found that EO provides a framework in which experience-based models of cognition can be embedded, which in turn allows for the behavior of the models to be optimized, in a similar fashion to the more standard parameter-fitting algorithms used in cognitive modeling. Furthermore, the methodology was tested using numerous cross-validation procedures and was shown to be able to fit

data at both the group and individual levels. The goal of the following simulations was to demonstrate that the information that EO is using when optimizing to a set of data is based on experiential factors (e.g., the time and place of data collection).

The basis of EO is that people's differential experience with language should be reflected in their lexical behavior (as we demonstrated above). Following on that premise, EO aims to optimize a model's fit to experimental measurements of lexical behavior by selecting the types of language materials that best reflect that experience. For example, Johns, Jones, and Mewhort (2019) showed that when optimizing the fit to lexical decision data from young versus old adults, EO selected young adult fiction books to account for the young adult lexical decision data, but selected more advanced fiction books to account for older adults' lexical decision data. Johns, Jones, and Mewhort demonstrated that models optimized in this fashion can achieve benchmark fits to data across a range of language behaviors.

EO is implemented by first assembling a large corpus of different text sources (in Johns, Jones, & Mewhort, 2019, the text sources used were fiction books, nonfiction books, young adult books, Wikipedia articles, and product descriptions from Amazon). These text sources are then split into smaller sections; typically, sections of 50,000 sentences (section size is a free parameter; however, using sections smaller than 50,000 sentences can lead to overfitting). Then a hill-climbing algorithm is used to iteratively select the best fitting section or sections.

For example, consider optimizing to a word familiarity dataset. The sections of the texts are first preprocessed into word frequency distributions. The first iteration in EO will select the first section, with this section being the one that has the best overall fit (e.g., assessed with a Pearson correlation coefficient) to a set of familiarity data. This section is then added into the optimized frequency distribution and removed from the search set (i.e., sampling without replacement). The second iteration will take all of the remaining sections and add each frequency distribution to the optimized frequency distribution. The resulting section that offers the best fit is added into the optimized frequency distribution and is removed from the search set. This process iterates until no section offers a meaningful improvement in fit. At the end of the search process, the selected corpus presents an optimized record of language experience in relation to the target data.

In the simulations that follow, each section given to EO will be based on the books written by an individual author whose place and time of birth are available. However, as can be seen in the book corpus descriptives shown in Table 1, there is significant variability in the word counts for individual authors. To solve the differences, we (a) excluded authors born before 1800, (b) excluded authors with fewer than 50,000 words in their assembled writings, and (c) represented the

writing of each of the remaining authors by a random sample of one million words from all of the words in their books. The sampling procedure in step (c) was used in order to ensure that each section had the same amount of lexical material. In total, 2,043 different sections (i.e., one section per author) were included in the search set. The result is a set of frequency distributions that include the equivalent amount of lexical information across authors, but differ in their distributional properties. In the following analyses, only familiarity rating and lexical decision norms were examined, because those data are available in much larger sample sizes than the category production data.

As an example of how EO operates, Fig. 9 includes the results of EO optimized to the familiarity data of Clark and Paivio (2004) and Gilhooly and Logie (1980). This figure shows that as new sections are selected, there is a corresponding increase in the fit of the optimized model to the data, with the Gilhooly and Logie dataset achieving a higher overall correlation. One reason for this discrepancy is that the language sources, as is visualized in Fig. 1, are skewed toward older sources, so older norms may receive a greater benefit using these language materials; however, item-level factors may be at play as well. Also shown, the most substantial increase in fit occurs at the outset of the search process, with smaller and diminishing improvements observed at later iterations. However, the final result is a fit to the data that is substantially better than fitting to the entire corpus. For example, the best correlation for the Clark and Paivio dataset was obtained for the new corpus, which had a correlation of $r = .67$, whereas the EO optimized frequency distribution was $r = .83$. For the Gilhooly and Logie dataset, the best corpus fit was

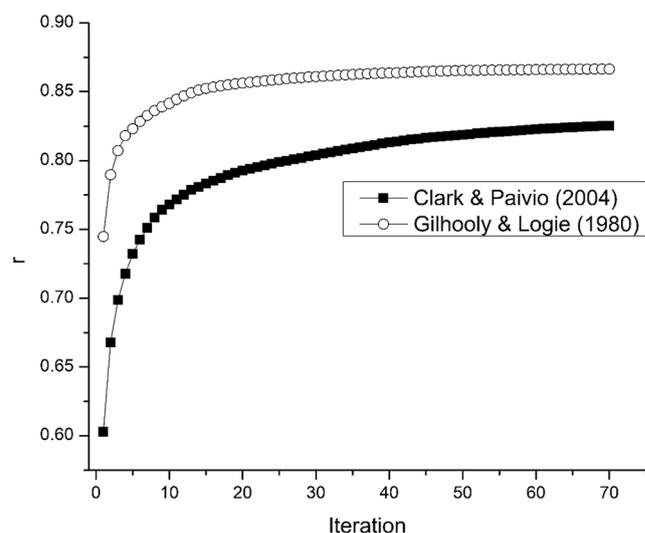


Fig. 9 Example of experiential optimization (EO) being applied to the word familiarity norms of Clark and Paivio (2004) and Gilhooly and Logie (1980). At each iteration, EO selects a language section, which causes an optimized frequency distribution to maximally increase its fit to a set of norms

$r = .78$, given by the UK corpus, whereas the EO optimized fit was an $r = .87$.

For the purposes of this article, the critical test of EO was whether it picked up on the differences in lexical behavior that are inherent in the empirical norms as a function of time and place. In the runs included in Fig. 9, the average date of birth of the authors sampled to the optimized corpus was greater for the more recent, Clark and Paivio (2004) norms ($M = 1948$) than for the older, Gilhooly and Logie (1980) norms ($M = 1943$). That difference in the right direction suggests that the optimization method can differentiate the influence of time on lexical behavior. In terms of author nationality, the method selected 57.5% UK authors when fitting the British Gilhooly and Logie norms, but only 14.84% UK authors when fitting the North American Clark and Paivio norms. This suggests that EO is also sensitive to the place where lexical norms were collected.

Johns, Jones, and Mewhort (2019) demonstrated that EO is sensitive to starting point, with different starting points causing different language sources to be selected. Thus, to ensure that the above findings are not anecdotal, 50 runs of EO were conducted for each set of word familiarity and lexical decision norms, with each run having a different random starting section. Then, the average date of birth of the authors selected and the percentage of authors who were born in the UK were recorded. If EO is picking up on the time and place of lexical behavior and matching that to the language corpus, then the average author date of birth and author place of birth selected by the method should reflect when and where the lexical data were collected.

An important point to take into consideration about the following simulations is that splitting language sources by the time and place of an author's birth is not the only way to categorize the likelihood of a group of participants' experiences with a certain author. Instead, time and place are broad and inclusive metrics, which might not map onto any single individual's language exposure particularly well. For example, Applebee (1992) found that 54% of sampled high school English courses in the USA taught *The Lord of the Flies* by William Golding, a British author born in 1911 (considerably older than the median date of birth in the author sample used here). Thus, the Golding section might be selected by EO more than other sections because of the ubiquity of his writing and the importance of his work to English literature. That is, his contributions would not be accounted for well by the time and place splits used in this article. However, that does not mean that the time and place splits do not have any meaning; indeed, the above simulations demonstrate that time and place seem to have a determinable impact on lexical behavior. Instead, it needs to be acknowledged that many factors go into any single individual's language exposure, which should be the focus for future research.

Figure 10 shows the fit of EO along with the best corpus fit (i.e., whether USA, UK, old, or new corpus) for the five sets of word familiarity norms and two sets of lexical decision norms. As is shown, EO yields an increased fit to lexical behavior for all ten datasets, over and above the otherwise best corpus fit. The results show the advantage of using EO to fit a model. However, the critical test will be to determine whether the optimization procedure was choosing sections that corresponded to the experience that participants likely had.

The top panel of Fig. 11 shows the average age of the sections selected to fit the older data (Gilhooly & Logie, 1980; Paivio et al., 1968; Stratton et al., 1975) and the newer data (Balota et al., 2007; Clark & Paivio, 2004; Keuleers et al., 2012; Stadthagen-Gonzalez & Davis, 2006). As can be seen, EO selected older authors to account for the older data and newer authors to account for the newer data. Even with this small sample size, the average date of birth selected for the new data was significantly greater than the average date of birth selected for the older data [$t(5) = 5.61, p = .002$]. Indeed, the correlation between publication date and average date of birth for the selected sections was $r(5) = .98, p < .001$. This demonstrates the EO is very sensitive to the effects of time on lexical behavior.

The bottom panel of Fig. 11 shows the percentages of authors selected who were born in the UK, relative to norms collected in the USA (Balota et al., 2007; Clark & Paivio, 2004; Paivio et al., 1968; Stratton et al., 1975) versus the UK (Gilhooly & Logie, 1980; Keuleers et al., 2012; Stadthagen-Gonzalez & Davis, 2006). This figure shows that EO tends to select language that corresponds to the place where the empirical data were collected, with the data

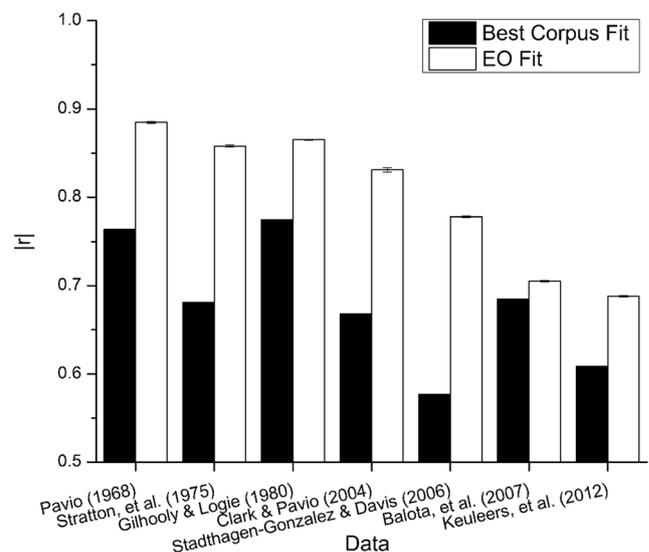


Fig. 10 Fits of the best targeted subcorpus (from the old, new, UK, or USA subcorpora; black bars) and the average optimized fits for 50 runs of experiential optimization (EO). Error bars show standard deviations of the means and demonstrate that there was little deviation in terms of fit when using EO from different starting points

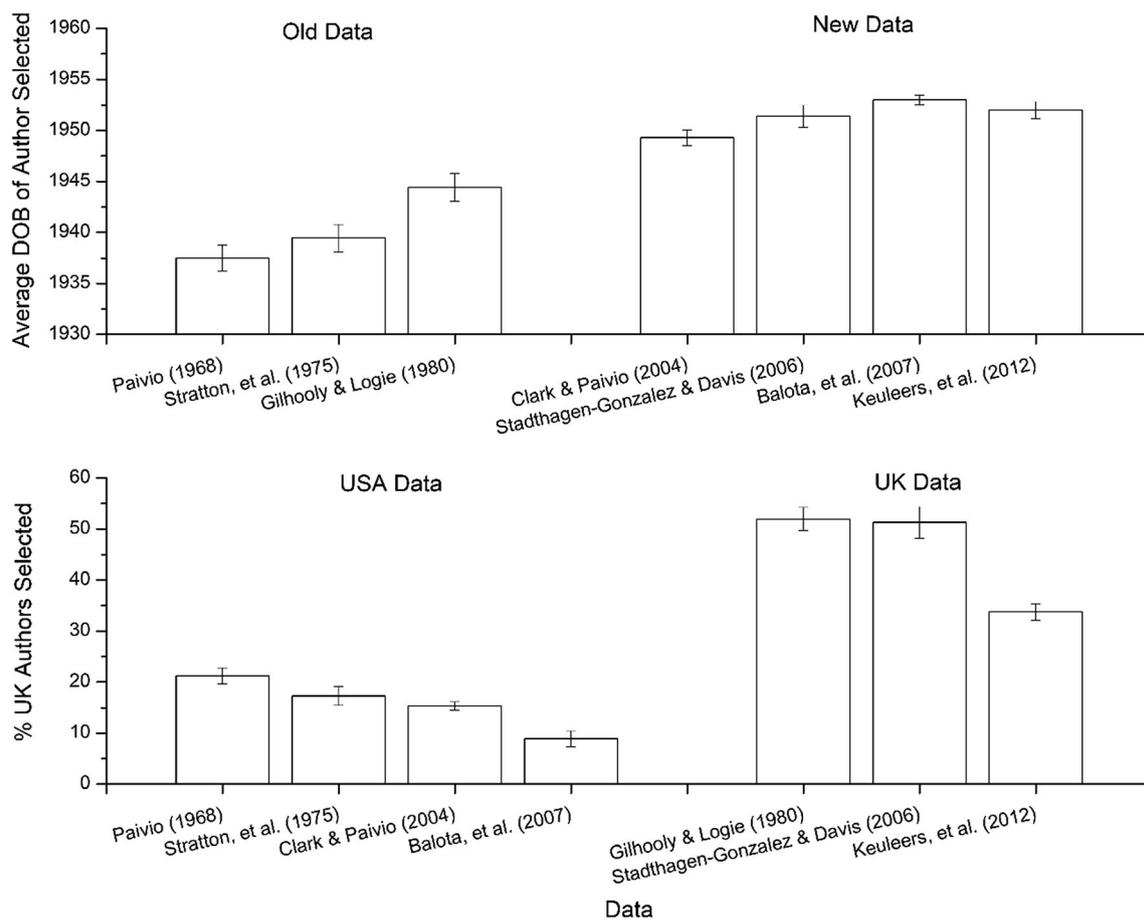


Fig. 11 Average dates of birth of the author sections selected (top panel) and the percentages of UK authors selected (bottom panel) when experiential optimization (EO) was applied to the assembled word

collected in the UK including many more sections written by UK-born authors [$t(5) = 5.194, p = .003$]. This is a particularly impressive result, considering that more than twice as many sections were provided by authors born in the USA ($n = 1,191$) than by authors born in the UK ($n = 565$), a result that suggests that EO is quite capable of selectively differentiating the place where lexical behavior was collected.

One unanswered question that the simulation in Fig. 11 raises is how much additional variance the addition of time- or place-matched language sections provide EO. To answer this question, the EO process was split into two runs of 20 iterations each (for a total of 40 iterations). On the first run, EO was constrained to choosing sections from authors born in either the USA or the UK. During the second run, the method was forced to only select sections from the other country. If the use of place-appropriate sections allows for EO to account for an increased amount of variance, then there should be a larger increase in fit when going from incongruent to congruent sections (e.g., fitting with USA and then UK sections when accounting for norms from the British Lexicon Project) then when going

familiarity and lexical decision norms. Error bars show standard deviations of the means. DOB = date of birth

from congruent to incongruent sections (e.g., fitting with UK and then USA sections when accounting for norms from the British Lexicon Project). As Fig. 9 shows, the expected increase in fit after 20 iterations is relatively small, but by comparing these two conditions it should allow for at least a better understanding of the amount of unique variance that using place-congruent sections of language provides when using EO. This was done for each set of the word familiarity and lexical decision norms, and the average increase in correlations was assessed across 50 simulations. The results of the simulation are displayed in Fig. 12. As is shown there, when the second run of EO uses congruent place sections, there is a consistent improvement in the method's ability to account for the experimental data. The result suggests that there is a unique place signature in language that the method is capable of identifying and leveraging.

Combined, these simulations demonstrate that EO optimizes a model to lexical data by tapping into the underlying lexical experience that groups of participants likely have, and not just by optimizing to noise in a dataset.

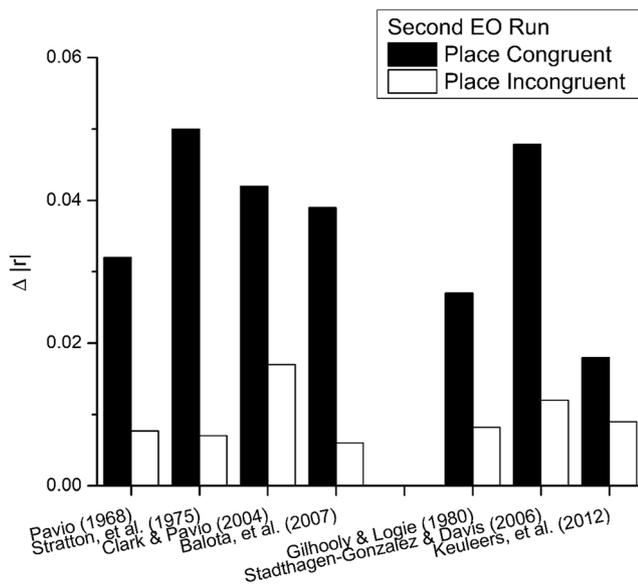


Fig. 12 Increases in correlations when experiential optimization (EO) sampled from a different place (incongruent condition) or the correct place (congruent condition) on the second run of the algorithm. As is shown, providing EO with sections of language consistent with the likely experience that a group of participants had yields a consistent benefit in the algorithm's ability to fit data

Supplementary material

To aid other researchers who are interested in the effects of experience on lexical behavior, this article has associated supplementary materials containing the various frequency distributions used in the simulations in this article.⁵ The word list for the frequency distributions was composed of the words from various familiarity datasets used here, the word list from the English Lexicon Project (Balota, et al., 2007), and the word list from the recently released word prevalence data of Brysbaert, Mandera, McCormick, and Keuleers (2019). This resulted in a final list of 81,186 words. The frequency distributions from the old, new, USA, and UK corpora were included in these materials, along with the frequency distribution from the overall book set described in Table 1. Additionally, the frequency distribution for each individual author, along with demographic characteristics recorded for each author, was also included, to allow other researchers to construct their own targeted or optimized frequency distributions.

General discussion

Our goal was to determine whether experiential differences in the language environment are reflected in people's lexical

behavior as a function of time and place. We call this the *selective-reading hypothesis*. To test the hypothesis, we collated a very large sample of books and tagged each book with its author's place and date of birth. An initial analysis of word use confirmed that books written by authors with the same place of birth (i.e., USA or UK) or the same time (i.e., generation) are more similar than books written by authors from different places of birth or from different generations. We evaluated the hypothesis against empirical norms for word familiarity, lexical decision, and semantic category production. Across the different sets of norms, models trained on time- and place-matched subcorpora offered the best fits to the experimental data. Taken together, the results support the conclusion that language models trained on time- and place-matched subcorpora can track the influence of language experience on people's lexical behavior and knowledge.

An additional goal was to use the assembled materials to test a new machine-learning framework for optimizing language models, called experiential optimization (Johns, Jones, & Mewhort, 2019). When optimizing a model to fit word familiarity and lexical decision data, EO chose sections of language that corresponded to the time and place in which the data were collected. This finding supports the conclusion that EO is capable of picking up on differences in language use as a function of place and time.

The underlying conceptualization and purpose of EO is to optimize language models by manipulating the experience that those models receive, instead of by varying internal cognitive parameters. That is, EO assumes that people from different places and times have different language experience and exhibit differences in lexical behavior as a consequence of that differential language experience, not of differences in language processing. The results of this article strengthen the argument for using EO when accounting for lexical behavior with cognitive models.

The work presented here shows that differences in language experience predict (and thus presumably influence) lexical organization and category knowledge. On those grounds, the work demonstrates how analyses of written text might be used to investigate the influence of experience on language processing. However, it points to a more fundamental issue for cognitive modeling and understanding the nested relationship between language and knowledge.

Typically, models of lexical organization and semantics are trained on a general (often convenient) corpus, and the capabilities of different learning schemes are assessed by relation to one another. This strategy is perfectly sensible—if you want to compare models, all other factors should be held constant. However, that strategy fails to appreciate the influence of language experience on the models. We have turned the typical strategy around, to ask how a single model behaves, depending on a change in the corpus. Just as changing models while holding the corpus constant produces different patterns of

⁵ The supplementary material can be found at http://www.btjohns.com/JJ_BR_M_Supplementary.rar.

behavior, our analysis shows that changing the corpus while holding the model constant produces different patterns of behavior. We chose time and place as a basis for the subcorpus distinction because this made for a broad and inclusive cut associated with variation in word meaning (e.g., *football*).

In his 1956 treatise, Herbert Simon pointed out that understanding cognition requires an examination and understanding of the organism, its environment, and the interaction of the two (Simon, 1956, 1969). Since that time, this mantra has been echoed and taken up in the discipline of ecological cognition (Todd & Gigerenzer, 2001, 2007) and elsewhere (e.g., Hills, Jones, & Todd, 2012; Johns, Jones, & Mewhort, 2019; Jones & Mewhort, 2007). Models of natural language processing have taken Simon's insight to heart and explained semantics as an emergent outcome arising from the operation of defined processing mechanisms (i.e., a model) and the language environment (i.e., a representative subcorpus of language experience). But, in large part, language experience has been treated agnostically, in deference to conducting a careful examination of the processing mechanisms. The work presented here shows that there is value in considering the language subcorpus just as carefully as the models themselves, and it offers a roadmap for engaging in a systematic and programmatic analysis of the influence that a person's language experience has on their lexical behavior.

References

- 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. <https://doi.org/10.1111/j.1467-9280.2006.01787.x>
- Applebee, A. N. (1992). Stability and change in the high-school canon. *English Journal*, *81*, 27–32.
- Baker, P. (2010). *Sociolinguistics and sub-corpus linguistics*. Edinburgh, UK: Edinburgh University Press.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *339*, 445–459. <https://doi.org/10.3758/BF03193014>
- Bartlett, F. C. (1928). An experiment upon repeated reproduction. *Journal of General Psychology*, *1*, 54–63.
- Bartlett, F. C. (1932). *Remembering: An experimental and social study*. Cambridge, UK: Cambridge University Press.
- Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*(3, Pt. 2), 1–46. <https://doi.org/10.1037/h0027577>
- Biber, D. (1993). Representativeness in sub-corpus design. *Literary and Linguistic Computing*, *8*, 243–257.
- Brysaert, M., Keuleers, E., & New, B. (2011). Assessing the usefulness of Google Books' word frequencies for psycholinguistic research on word processing. *Frontiers in Psychology*, *2*, 27.
- Brysaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*, 45–50. <https://doi.org/10.1177/0963721417727521>
- Brysaert, M., Mandera, P., McCormick, S.F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*, 467–479. <https://doi.org/10.3758/s13428-018-1077-9>
- Brysaert, 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. <https://doi.org/10.3758/BRM.41.4.977>
- Cai, Q., & Brysaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*, e10729:1–8. <https://doi.org/10.1371/journal.pone.0010729>
- Chubala, C. M., Johns, B. T., Jamieson, R. K., & Mewhort, D. J. K. (2016). Applying an exemplar model to an implicit rule-learning task: Implicit learning of semantic structure. *Quarterly Journal of Experimental Psychology*, *69*, 1049–1055.
- Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 371–383. <https://doi.org/10.3758/BF03195584>
- Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*, 395–427. <https://doi.org/10.3758/BF03201693>
- Green, C. D., Feinerer, I., & Burman, J. T. (2013). Beyond the schools of psychology 1: A digital analysis of *Psychological Review*, 1894–1903. *Journal of the History of the Behavioral Sciences*, *49*, 167–189.
- Green, C. D., Feinerer, I., & Burman, J. T. (2015). Searching for the structure of early American psychology: Networking *Psychological Review*, 1894–1908. *History of Psychology*, *18*, 15–31.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*, 211–244. <https://doi.org/10.1037/0033-295X.114.2.211>
- Hampton, J. A., & Gardiner, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, *74*, 491–516.
- Herdağdelen, A., & Marelli, M. (2017). Social media and language processing: How Facebook and Twitter provide the best frequency estimates for studying word recognition. *Cognitive Science*, *41*, 976–995. <https://doi.org/10.1111/cogs.12392>
- Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*, 431–440. <https://doi.org/10.1037/a0027373>
- Johns, B. T. (2019). Mining a crowdsourced dictionary to understand consistency and preference in word meanings. *Frontiers in Psychology*, *10*, 268. <https://doi.org/10.3389/fpsyg.2019.00268>
- Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *Journal of Acoustical Society of America*, *132*, EL74–EL80.
- Johns, B. T., & Jamieson, R. K. (2018). A large-scale analysis of variance in written language. *Cognitive Science*, *42*, 1360–1374. <https://doi.org/10.1111/cogs.12583>
- Johns, B. T., & Jones, M. N. (2015). Generating structure from experience: A retrieval-based model of language processing. *Canadian Journal of Experimental Psychology*, *69*, 233–251.
- Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2019). Using experiential optimization to build lexical representations. *Psychonomic Bulletin & Review*, *26*, 103–126. <https://doi.org/10.3758/s13423-018-1501-2>
- Johns, B. T., Mewhort, D. J. K., & Jones, M. N. (2019). The role of negative information in distributional semantic learning. *Cognitive Science*, *43*, e12730. <https://doi.org/10.1111/cogs.12730>

- Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The role of semantic diversity in word recognition across aging and bilingualism. *Frontiers in Psychology*, *7*, 703:1–11. <https://doi.org/10.3389/fpsyg.2016.00703>
- Jones, M. N. (2017). *Developing cognitive theory by mining large-scale naturalistic data*. In M. N. Jones (Ed.), *Big data in cognitive science*. New York, NY: Taylor & Francis.
- Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizational principle of the lexicon. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 67, 239–283). San Diego, CA: Elsevier Academic Press.
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, *66*, 115–124. <https://doi.org/10.1037/a0026727>
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1–37.
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643–650. <https://doi.org/10.3758/BRM.42.3.643>
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*, 287–304. <https://doi.org/10.3758/s13428-011-0118-4>
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240. <https://doi.org/10.1037/0033-295X.104.2.211>
- Mandera, P., Keuleers, E., Wodniecka, Z., & Brysbaert, M. (2015). SUBTLEX-PL: Subtitle-based word frequency estimates for Polish. *Behavior Research Methods*, *47*, 471–483. <https://doi.org/10.3758/s13428-014-0489-4>
- Paivio A. (1974). [Imagery and familiarity ratings for 2,448 words] (Unpublished norms). London, ON: University of Western Ontario, Department of Psychology.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1, Pt. 2), 1–25. <https://doi.org/10.1037/h0025327>
- Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, *6*, 623–641.
- Recchia, G., Sahlgren, M., Kanerva, P., & Jones, M. N. (2015). Encoding sequential information in semantic space models: Comparing holographic reduced representation and random permutation. *Computational Intelligence and Neuroscience*, *2015*, 986574. <https://doi.org/10.1155/2015/986574>
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9)
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.
- Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*, 598–605. <https://doi.org/10.3758/BF03193891>
- Stratton, R. P., Jacobus, K. A., & Brinley, B. (1975). Age-of-acquisition, imagery, familiarity and meaningfulness norms for 543 words. *Behavior Research Methods & Instrumentation*, *7*, 1–6. <https://doi.org/10.3758/BF03201278>
- Taler, V., Johns, B. T., & Jones, M. N. (2019). A large scale semantic analysis of verbal fluency across the aging spectrum: Data from the Canadian Longitudinal Study on Aging. *Journals of Gerontology B: Psychological Sciences*. Advance online publication. <https://doi.org/10.1093/geronb/gbz003>
- Todd, P. M., & Gigerenzer, G. (2001). Shepard's mirrors or Simon's scissors? Commentary on R. Shepard, "Perceptual-cognitive universals as reflections of the world." *Behavioral and Brain Sciences*, *24*, 704–705. <https://doi.org/10.1017/S0140525X01650088>
- Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, *16*, 167–171.
- Tremblay, M., & Vézina, H. (2000). New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *American Journal of Human Genetics*, *66*, 651–658.
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*, 289–335. <https://doi.org/10.1016/j.jml.2003.10.003>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.