Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale

Scalable Cognitive Modelling: Putting Simon's (1969) Ant Back on the Beach

Brendan T. Johns, Randall K. Jamieson, and Michael N. Jones

Online First Publication, April 10, 2023. https://dx.doi.org/10.1037/cep0000306

Canadian Journal of Experimental Psychology /
Revue canadienne de psychologie expérimentale

ISSN: 1196-1961

https://doi.org/10.1037/cep0000306

Scalable Cognitive Modelling: Putting Simon's (1969) Ant Back on the Beach

Brendan T. Johns[1], Randall K. Jamieson[2], and Michael N. Jones[3]
[1] Department of Psychology, McGill University
[2] Department of Psychology, University of Manitoba
[3] Department of Psychological and Brain Sciences, Indiana University

A classic goal in cognitive modelling is the integration of process and representation to form complete theories of human cognition (Estes, 1955). This goal is best encapsulated by the seminal work of Simon (1969) who proposed the parable of the ant to describe the importance of understanding the environment that a person is embedded within when constructing theories of cognition. However, typical assumptions in accounting for the role of representation in computational cognitive models do not accurately represent the contents of memory (Johns & Jones, 2010). Recent developments in machine learning and big data approaches to cognition, referred to as scaled cognitive modelling here, offer a potential solution to the integration of process and representation. This article will review standard practices and assumptions that take place in cognitive modelling, how new big data and machine learning approaches modify these practices, and the directions that future research should take. The goal of the article is to ground big data and machine learning approaches that are emerging in the cognitive sciences within classic cognitive theoretical principles to provide a constructive pathway towards the integration of cognitive theory with advanced computational methodology.

Public Significance Statement
Computational modelling has played a central role in the development of theory in cognitive psychology. Recently, machine learning and big data approaches to understanding cognition have become increasingly popular. This article reviews standard approaches in computational cognitive modelling and specifies how new advanced computational approaches can be used to generate new research pathways in the cognitive sciences.

Keywords: cognitive modelling, machine learning, big data, lexical semantics, distributional modelling

Two principal strategies have been used to build knowledge into computational systems. The first is to program sophistication into the system itself so that its behaviour reflects built-in (i.e., innate) knowledge, such as early work in models of semantic memory in psychology (e.g., Collins & Quillian, 1969). Examples of that approach include expert systems in artificial intelligence research (Jackson, 1986). The other strategy is to provide the system with a sufficient store of data in a relevant domain so that sufficient representations for the problem can be learned by the system. Examples of that approach include modern machine learning (e.g., LeCun et al., 2015). Although both approaches can solve similar problems, they nevertheless represent very different philosophies about the nature of knowledge and intelligence when considered as models of human cognition.

By embedding a model with hand-coded knowledge, the system is prepared for the world in which it lives and survives by that "intelligent design." That is, a system of this type's ability to behave within the environment that it is embedded in is dictated by its programming, often informed by domain expertise from human experts. For example, an automated robot in a car manufacturing factory only needs to know how to correctly weld, paint, or assemble. For these types of applications, there is no need for these systems to adapt. Instead, the emphasis is placed on optimizing the system's behaviour to perform a specific and anticipated task as perfectly as possible.

In contrast, machine learning approaches to cognition and software development acknowledge the noise and uncertainty of the world. In many cases, it is impossible to forecast all possibilities and contingencies that a system will face when inserted into the real world. For example, in automated speech or handwriting recognition, there is significant variability in the information that the systems will

Editor's Note. Debra Titone served as the action editor for this article.—DT

Brendan T. Johns https://orcid.org/0000-0001-6152-4914
Randall K. Jamieson is the previous editor of this journal. This research was supported by the Natural Science and Engineering Research Council of Canada Discovery Grant RGPIN-2020-04727 to Brendan T. Johns.
Correspondence concerning this article should be addressed to Brendan T. Johns, Department of Psychology, McGill University, 2001 McGill College Avenue, Montreal, QC H3A 1G1, Canada. Email: brendan.johns@mcgill.ca

encounter (Goldinger, 1998). Thus, instead of building in abstracted knowledge about the structural properties of a language (e.g., phonemes and graphemes), machine learning approaches train models on noisy, varied, and diverse information sources in the hope that the system can outmanoeuvre all of the unforeseen circumstances and variations that it might come across when applied to real input (e.g., thousands or millions of samples of written and spoken language; Graves et al., 2013; Graves & Schmidhuber, 2008).

Similar arguments have occupied debates about the nature of knowledge across the history of the philosophy of mind, psychology, and linguistics (e.g., Chomsky, 1991; Jackendoff, 1992; Pinker, 1994; Quine, 1960). In those circumstances, the division is framed in terms of nature versus nurture; a debate still alive in both the psychological and biological sciences (e.g., Gould, 1996; Herrnstein & Murray, 2010). Indeed, such arguments are now taking place in the cognitive sciences in response to the development of machine learning approaches to cognition (e.g., Günther et al., 2019; Johns, 2022a; Kumar, 2021; Landauer & Dumais, 1997; Thompson et al., 2020). With the entry of computational modelling to that debate, we see real opportunity to advance our understanding of and even potentially resolve these issues.

In this article, we make the case that nature might play a role in setting the conditions and parameters of cognition, but that nurture and learning play a dominant role in how behaviour emerges against the often unpleasant but necessary motivations of living. More critically, it will be pointed out that the toy examples of cognitive behaviour and learning that have been and generally continue to be explored in the psychological laboratory have, in combination with the advent of our digital lives in the 21st century, put cognitive researchers in a watershed moment. Indeed, it is now possible to take the careful lessons learned from laboratory science, and the precise cognitive models derived from it, into the broader and scaled-up world of cultural knowledge. Doing so will bring psychology out of the laboratory and deliver on the promise from long ago to situate psychology in a primary place amongst the biological and information sciences (see Hebb, 1958; Mewhort, 1990).

Hebb (1958) discussed the proper role of psychology. He likened psychology to Alice in Wonderland and how our discipline needed to find its place in the biological sciences. That treatise led to the modern shape of psychology, particularly the influence of cognitive neuroscience in modern psychological science. Mewhort (1990), an intellectual grandchild of Hebb's, built on Hebb's arguments to discuss psychology's place in the information sciences and proposed that psychologists had an important role to play in the design and implementation of artificial intelligence and what has become known as cognitive computing, machine learning, and data science. Mewhort discussed how artificial neural networks at the time provided excellent accounts of local laboratory tasks (e.g., the interactive activation model for understanding the word superiority effect; McClelland & Rumelhart, 1981), but that those systems did not scale up to explain behaviour beyond the contrived laboratory examples on which they were developed (Feldman-Stewart & Mewhort, 1994). Yet, he expressed a belief that the limitation reflected the failure of psychologists to think at the scales of the problems they wanted to solve. Accordingly, he encouraged the discipline to think about "psychology at scale" so as to bring psychologist's unique perspectives and contributions to bear on the goals of technological invention and innovation.

The motivations of accounting for behaviour at scale have a rich history in the cognitive sciences. Indeed, such notions were at the heart of Simon's (1969) vision for the future of the discipline. To underscore the importance of environmental structure on intelligence and behaviour, Simon proposed the parable of the ant. In this parable, he describes the types of theories that one may entertain about an ant's behaviour when observing it walking along a beach. If one only considers the ant's path, without also considering the environment in which that path lay, one might ascribe the ant's meandering path to a complex internal mechanism. However, if one also considers the environmental obstacles in the ant's way (i.e., a branch here and a stone there), one would realize that the ant was using simple internal mechanisms in reaction to a complex environment (i.e., turn left at the branch and right at the stone in order to go in a straight line). That is, from Simon's point of view, constructing a complete theory of behaviour and cognition requires an awareness and integration of the tasks at hand, the organism faced with the task, and the environment in which the task must be accomplished.

Models of cognition have both processing and representation assumptions, with both aspects of a model being interdependent (Kanerva, 2009). For example, if one wanted to construct a model of episodic memory, one would need to specify exactly what is contained in an episodic memory trace and also the process by which those traces are retrieved from memory. However, this interdependency has been ignored in the implementations of most cognitive models. In particular, the majority of cognitive models of episodic memory have utilized randomized representations or artificially constructed representations of memory content, especially when that content refers to semantic information (Johns & Jones, 2010). That is, most cognitive models have assumed away variance in the representations learned from the environment, in order to focus on determining the underlying processing mechanisms of cognition. While important and fundamental information about memory processes has been attained using this strategy, for cognitive models to increase in plausibility and power, it is necessary to also consider the content that is contained in the representations of memory and language.

People learn from experience within the environments that they occupy and are embedded in, from the social to the perceptual to the linguistic (Tiv et al., 2022). This presents a challenge to cognitive models seeking to integrate scaled and realistic representational assumptions into their architecture. However, the rise of big data sources and corresponding machine learning algorithms that can learn from these sources has allowed for remarkable recent progress. In the cognitive sciences, the area of study where the impact of big data and machine learning has been most pronounced is the study of lexical semantics (see Günther et al., 2019; Kumar, 2021, for recent reviews), beginning with the classic latent semantic analysis (LSA) model of Landauer and Dumais (1997).

The work of Landauer and Dumais (1997) was prompted by a central problem in the philosophy of mind, dating back to Plato, about whether knowledge acquisition is based upon innate constraints or from statistical learning of environmental structure (to put it into modern terms). LSA demonstrated that a relatively simple learning mechanism, based upon techniques derived in the field of information retrieval in databases (Deerwester et al., 1990), and trained with a sufficient amount of natural language, can provide impressive fits to tests of human semantic knowledge (e.g., it performed at about the same level as an English a second language user on a synonym test). This result has prompted a broad range of model development exploring different mechanisms to extract semantic representations from large collections of text (e.g., Griffiths et al., 2007; Jamieson et al., 2018; Jones & Mewhort, 2007; Landauer & Dumais, 1997; Mikolov et al., 2013).

An even earlier example of the importance of big data in the psychological and cognitive sciences is given by the extraction of word frequency values by Kucera and Francis (1967). These values were collected by measuring the frequency of words from a sample of approximately 1 million words from a corpus assembled from a variety of sources, such as newspaper articles and fiction novels. Assembling these values was prompted by early results in experimental psycholinguistics demonstrating that words that appeared more often were processed faster (e.g, Broadbent, 1967). Subsequently, word frequency has been established as a central methodological and theoretical concept in the study of lexical organization (see Brysbaert et al., 2018, for a review).

Kučera and Francis (1967) faced significant challenges assembling their materials given the technological limitations of the time, which explains the longevity of the impact of these norms (Brysbaert & New, 2009). Modern researchers do not face these same limitations. Recently, there have been a number of additional word frequency sets from a diverse number of sources that have been collected and disseminated, such as television and movie subtitles (Brysbaert & New, 2009), newspaper articles (Davies, 2009), online encyclopaedias (Shaoul & Westbury, 2010), fiction books (Johns, Dye, et al., 2020; Johns & Jamieson, 2019), social media (Herdağdelen & Marelli, 2017), and online forums (Johns, 2019, 2021a). This increase in types of lexical materials tracks cultural developments in the use of language across different technological mediums. These examples demonstrate the initial promise of using new technological developments in the computational sciences to drive developments in the psychological and cognitive sciences. However, for these collections of texts to be useful in the development of cognitive theory, they also need to be used by cognitive models, a significant challenge. However, before discussing this possibility, the interdependency of representation and process in cognition will be explored.

Processing and Representation in Cognitive Modelling

To illustrate the importance of both representation and processing assumptions in the development of cognitive theory, consider the following scenario: a student's advisor tells them that "you want to be where the science will be, not where it is." The student appreciates this advice, but they have a nagging feeling that they had heard something similar before. While searching the internet to determine the cause of their feeling of familiarity, the student realized that it was based on a famous quote from the hockey player Wayne Gretzky, who stated that "I skate to where the puck is going to be, not to where it has been."

The student's ability to identify that their advisor's advice was similar to a previously heard quote is dependent on two main components of memory and language. First, it is necessary to have a comparative process where the structure of the current episodic experience is compared against past experiences (i.e., retrieval). Second, there needs to be an underlying representation of the meaning and not just the specific words of those experiences so that the process can determine the similarity of the present context to past experience (i.e., knowledge). That is, both a processing mechanism and a representation type need to be proposed in order to successfully account for student's recognition memory performance.

A standard example of a cognitive model with well-defined processing and representation assumptions is given with the classic MINERVA 2 model of Hintzman (1986, 1988). MINERVA 2 is a multitrace memory model, employed to explain a variety of phenomena across memory and language since its inception (e.g., Arndt & Hirshman, 1998; Chubala et al., 2016; Goldinger, 1998; Jamieson et al., 2012; Jamieson & Mewhort, 2009, 2010, 2011; Johns, Jamieson, et al., 2020; Johns & Jones, 2012, 2015; Kwantes, 2005; Thiessen & Pavlik, 2013; see Jamieson et al., 2022, for a recent review). In this model, each stimulus that a model encounters is stored as a separate trace in memory. For example, when MINERVA 2 is applied to understand list memory performance, the representation of each word contained in a list is stored in a different location in a memory store. Decision in the model is based upon a transformation of the summed similarity between the representation of a presented probe to the stored representations in memory. If the summed similarity value (the probe's global familiarity) over all traces exceeds a set criterion value, then the probe is accepted; otherwise, it is rejected.

MINERVA 2's calculation of a probe's global familiarity is dependent on the representation of both probe and memory items. MINERVA 2 uses randomly generated feature vectors to define these representations. Specifically, each item is represented with a vector (typically of a low dimensionality), containing values of $\{0, 1, -1\}$, where 1 and -1 represent having a feature or not, and 0 represents lack of encoding (controlled with a free parameter designed to mimic cognitive factors such as learning rate or attentional variability). If the probe occurred in a study list, its global familiarity tends to be high enough to accept (i.e., it will exceed the set criterion), whereas if the probe did not occur in the study list, it tends to have a lower global familiarity and can be rejected (i.e., it will be below the set criterion). Likewise, if the probe has small amounts of similarity to several studied items, it might also be accepted due to a relatively high global familiarity; a feature that allows for the model to account for phenomena such as false recognition (Arndt & Hirshman, 1998).

However, the use of randomly generated representations ignores a central part of the architecture of a model of memory, namely the content of the stimuli that is being used in an experiment. The use of randomly generated representations is not just isolated to MINERVA 2, but indeed has been a standard assumption in computational models of memory and language (e.g., Cree et al., 1999; Dennis & Humphreys, 2001; Murdock, 1982; Shiffrin & Steyvers, 1997, to name a few). This is not a failure of these models but is a reflection of the difficulty in accounting for representational complexity due to technological limitations, specifically in building realistic representation types that map onto item-level properties (e.g., the meaning of words).

In order to get a handle on the impact that the user of randomly generated representations has on model assumptions and performance, Johns and Jones (2010) compared the similarity structures that different random representation types constructed and compared them to the similarity structures that different vector-based lexical semantic models (mostly distributional models; see below for more details on this model type) produced. This was done due to the fact that researchers typically use words in episodic memory experiments, and models of recognition memory are typically dependent on computations of similarity (typically focused on word meanings) between a probe word and a list of studied words. It was found that the similarity distributions produced by the semantic measures were systematically different than the representations derived by the random representations, suggesting that the assumptions that the models were making about the nature of representation were incorrect.

Furthermore, Johns and Jones (2010) demonstrated that when the MINERVA 2 model was used to examine false memory effects (similar to efforts of Arndt & Hirshman, 1998), the model was unable to account for the correct pattern of data when given a representation that contained semantic information about words but could account for the data when fit with randomized representations. This suggests that the use of random representations also allowed for increased flexibility in accounting for data and does not just signal an incorrect assumption about the nature of cognition. That is, by being able to manipulate both the process that produces a behaviour and the information structures contained in memory representations, a modeller has a greater level of flexibility in model development than relying upon a fixed representation type.

An early example of the power of constructing realistic representations of stimuli is given by Nosofsky (1986) when modelling the interaction between attention and identification in categorization. In the study, two participants conducted a perceptual identification task where they were presented with circles of four sizes and lines with four different orientations, and the participants had to identify each unique object with a 4×4 button grid. Those identification data were used to generate a confusion matrix that recorded the number of times that each button was pressed when a participant was presented with each stimulus. This matrix was then reduced with multidimensional scaling to construct a two-dimensional representation of each stimulus. It was found that using this derived representation as the underlying basis of a model of categorization allowed for the model to achieve excellent fits to behavioural data, even though the representations were not derived from a categorization task. This result demonstrates the importance of accounting for both processing and representational components of a cognitive model and the power that comes with it.

However, the stimuli that Nosofsky (1986) used were artificially designed to be objectively defined along articulate and known stimulus dimensions (although see Nosofsky et al., 2020; Sanders & Nosofsky, 2020, for similar work using model-based visual representations of natural images). How to integrate representations of more complex stimulus types, such as words, into cognitive models offers a considerable challenge to cognitive modellers. The meaning of words is acquired through millions of episodic experiences with language, and the dimensions of meaning are both difficult to identify and latent in people's semantic interpretation (Brysbaert et al., 2016; Hollis & Westbury, 2016). Until recently, building cognitive representations for words at scale was impossible due to a variety of limitations, both in computational power available to researchers as well as the limited availability of scaled training materials. However, developments in cognitive theory based around scaled cognition (such as the classic LSA model of Landauer & Dumais, 1997, and the models inspired by it) and the continual evolution of computational hardware and the construction of varied and large linguistic corpora have made generating realistic representation types ever more possible.

The Simplification Assumption in Cognitive Modelling

The rationale for using randomly generated representations in cognitive modelling is best described by the simplification assumption (McClelland, 2009; Shiffrin, 2010; see Johns et al., 2017, for a more in-depth discussion of these issues). The goal of this assumption is to allow for a modelling exercise to focus on the important aspects of a model's performance while assuming that other components of the model can be held constant with minimal theoretical or computational commitments. As justification for this assumption, McClelland (2009, p. 18) states, "The more detail we incorporate, the harder the model is to understand."

A good example of the use and power of the simplification assumption is given by the semantic cognition model of Rogers and McClelland (2004). This model utilizes a connectionist architecture based on the previous work of Rumelhart (1990) and Rumelhart and Todd (1993), with the goal of the model being to examine the dynamics of the acquisition of meaning from a developmental point of view. The model is experientially dependent and requires training materials. To train the model, the model was given propositions derived from semantic networks based on the early work of Collins and Quillian (1969). This network contained propositions about the nature of plants and animals (e.g., *canary is a animal* or *daisy is yellow*). In the network from Collins and Quillian (1969), the network contained a total of 84 propositions to describe all of the relations. Additionally, they used a larger network that contained 220 propositions.

In the account of Rogers and McClelland (2004), the representational complexity of the model is kept constant. That is, the training materials that the model is presented with are structured such that the learning dynamics can be well understood while ignoring the actual complexity that is contained in the natural language environment. As an example of the differences in complexity of the propositions in the training materials of Rogers and McClelland (2004), consider Table 1. This table contains the propositions derived from the semantic network of Collins and Quillian (1969) for the words *canary* and *pine*, as well as sentences for these words attained from young adult fiction novels used previously by Johns and Jamieson (2018, 2019) and Johns, Dye, et al. (2020). This table shows that the two language sources differ significantly in the complexity of the information that is contained about the meaning of these words, with the propositions being much more straightforward in terms of elucidating the meaning of the two words, and also the type of language used, with the sentences from the natural language corpora being much more figurative in meaning. That is, sentences contained in Table 1 are qualitatively different from each other and models seeking to learn the meaning of words would face a much harder time deriving the meaning of canary and pine when trained on a noisy and natural corpus compared to the articulate and contrived propositions from Collins and Quillian's (1969) semantic network.

The model of Rogers and McClelland (2004) is one designed to explore the development of lexical semantics, and the evaluation of the model is not focused on whether it can learn the structure of the training materials (many different mechanisms could learn the underlying structure of such simple materials)—instead, the emphasis is placed on understanding developmental trajectories in semantic memory. Given that human children are being exposed to much more complex information from their environment compared to the network from Collins and Quillian (1969), these materials do not accurately reflect the learning task that a human child faces.

Table 1
Examples of Propositions Derived From the Semantic Network of Collins and Quillian (1969) for the Words Canary and Pine and Sentences From Young Adult Novels for These Same Words

Word	Proposition	Sentence
Canary	Canary is a bird	You both look like the cat that swallowed the canary.
	Canary is a animal	She had said goodbye to her canary.
	Canary is a living_thing	Our canary has stopped singing.
	Canary can grow	He is not imprisoned like a canary.
	Canary is living	There was actually a canary inside the birdcage.
	Canary can move	He is my pet canary.
	Canary has skin	I'm not a canary.
	Canary has feathers	Jack's just the canary in the coal mine.
	Canary can fly	His canary had fallen into a terrified cheeping.
	Canary has wings	Much better than a canary that won't sing.
	Canary is yellow	I'm a freaking giant canary but thanks anyway.
	Canary can sing	I got a canary down there.
Pine	Pine is a tree	They went to sit down under the pine trees.
	Pine is green	I got a cargo of yellow pine from Labrador.
	Pine is a plant	Going to see a shipment of pine that just came in.
	Pine is a living_thing	There were also a lot of pine and aspen trees.
	Pine can grow	The sound echoed through the pine trees.
	Pine is living	Your shout could fell a pine tree.
	Pine has roots	The warrior was as tall as a pine tree.
	Pine has bark	More like red bull and pine-scented deodorant.
	Pine is big	The entire room smelled like sweet pine needles.
	Pine has branches	She led them toward the big pine tree.

In order to gain an understanding of the task difficulty that distributional models face in extracting meaning from the natural language environment, Johns et al. (2017) used a recently developed optimization procedure for distributional models (entitled experiential optimization [EO]; Johns, Jones, et al., 2019; see below for more detail) to determine how much experience, in terms of number of sentences, the bound encoding of the aggregate language environment (BEAGLE) model of semantics (Jones & Mewhort, 2007) required to learn the structure of the semantic network of Collins and Quillian (1969). It was found that BEAGLE required 150,000 curated natural language sentences to reach the performance of the same model trained on 100 propositions from the Collins and Quillian (1969) network. This demonstrates that even when training materials from the natural language environment are optimized, there is significant noise in corpus-based sources that requires a great deal of experience for a model to extract stable, let alone empirically matched, word meanings. For researchers interested in examining learning trajectories in language development, the greater complexity of the natural language environment presents a significantly greater challenge for distributional models (however, see Asr et al., 2016; Braginsky et al., 2019; Huebner & Willits, 2018, for some examples of this type of research).

Given the inherent complexity of the natural language environment, the simplification assumption has been used across both empirical and theoretical research in psycholinguistics and episodic memory research to isolate important theoretical questions and has proven to be of great importance in cognitive science. Empirically, tasks such as artificial grammar learning (Reber, 1976) or artificial language learning (e.g., Gómez & Gerken, 2000) have been used to examine language acquisition using limited timeframes in order to gain insight into the learning mechanisms that are at play in statistical language acquisition across the lifespan. Furthermore, the use of simplified training has been popular to train and evaluate models of language since the pioneering work of Elman (1990), where simplified training materials were used in the development of recurrent neural networks. Similar strategies have been used across a number of different areas of language processing, for example, in the modelling of syntactic priming (Chang et al., 2006) and event knowledge (Elman & McRae, 2019). Indeed, artificial languages have been used to evaluate differences in the mechanisms that different distributional models use (Asr & Jones, 2017; Jamieson et al., 2018).

Research done utilizing the simplification assumption has proven to be fundamental in the development of cognitive theory. However, there are multiple reasons for this success. One reason is due to the description of the simplification assumption as put forth by McClelland (2009)—that by isolating processing mechanisms from representational complexities, a better understanding of the capabilities of the processing mechanisms can be attained, as displayed by Elman (1990), Rogers and McClelland (2004), and the various episodic memory models described previously. By assuming away representational complexity, it is easier to gain an understanding of the successes and failures of proposed processing mechanisms. As Mewhort (1990) pointed out, however, the benefits of simplification can represent themselves as costs when the task of scaling theories up to real-world complexity comes to roost.

Another reason for the simplification assumption being embraced by various researchers is due to the limitations in the training materials that have been available to researchers up until recent times. The availability of large-scale training materials that can be used to train cognitive models has only recently been developed. For example, the Touchstone Applied Sciences Associates corpus of Landauer and Dumais (1997) was one of the first major corpora available of sufficient quality to develop computational models with. Researchers now have a wide variety of different language materials to train models (see Johns, Jones, et al., 2019, for a framework utilizing a wide range of different types of language materials) due to the digitization of diverse collections of text.

However, the cognitive system that models of language and memory are aiming to solve are those that require lifelong experience to explain, as the impact of accumulated language experience has differential impacts on language and memory across the lifespan (Dubossarsky et al., 2017; Qiu & Johns, 2020; Ramscar et al., 2014; Taler et al., 2020; Wulff et al., 2019; see Ramscar, 2022, for a review). That is, the complexities contained in the natural language environment are significantly greater than are assumed by the simplified training materials. For computational cognitive science to continue to evolve and innovate as a field, it will be increasingly necessary to use more complex training materials to reduce the number of assumptions that underlie a model, especially in terms of representational assumptions.

Scaled Approaches to Cognitive Modelling

As stated previously, the study of lexical organization was one of the first areas of cognitive science to embrace large-scale analyses of the language environment in order to understand the connection between the language environment that people are embedded in and corresponding lexical behaviour. Kucera and Francis (1967) published the first widely available word frequency norms, attained from a sample of 1 million words across different types of texts. The use of these norms has been widespread across the cognitive sciences over the last 50 years, with Kucera and Francis (1967) being cited over 8,000 times since publication, demonstrating the importance of word frequency to the field of psychology—a corollary of the pervasive and apparent impacts of word frequency on lexical behaviour (see Brysbaert et al., 2018, for a review) and memory performance (e.g., Glanzer & Adams, 1985, 1990).

Word frequency is an important measure used across the cognitive sciences, from being a central component of theoretical accounts of word recognition and lexical organization (e.g., Coltheart et al., 2001; Goldinger, 1998; Murray & Forster, 2004; Norris, 2006) to methodological considerations in stimulus selection (e.g., Brysbaert & New, 2009). The use of word frequency in cognitive models acknowledges the role that accumulated experience with the language environment has on the language and memory processing system. In terms of lexical organization, the use of frequency as an organizational principle acknowledges that words that occur more often in the language environment should be stored more strongly, or should be more easily accessible, in memory. By endowing models with word frequency values, it allows for lexical storage models to simulate the impact of accumulated language experience.

However, scaled approaches to lexical organization have recently been questioning the primacy of word frequency in accounting for the organization of the mental lexicon, with these theories arising from corpus-based analyses. The first major measure to challenge word frequency was entitled contextual diversity and was first proposed by Adelman et al. (2006; although see McDonald & Shillcock, 2001, for a similar, earlier proposal). That proposal suggests that the strength of a word should be updated not on each occurrence of a word, but each new context that a word occurs in. To operationalize a context, natural linguistic markers were used (e.g., document, article, chapter, book). It has been found that a context count systematically outperforms word frequency across multiple large data sets of lexical behaviour (Adelman & Brown, 2008; Adelman et al., 2006; Brysbaert & New, 2009; Johns, 2021a, 2022b; Johns, Dye, et al., 2020; Johns et al., in press; Johns & Jones, 2022; Jones et al., 2012; Senaldi et al., 2022; see Jones et al., 2017 and Caldwell-Harris, 2021, for a review).

The success of contextual diversity measures of lexical organization signals the power of scaled approaches to cognitive modelling, as the availability of large text sources allowed for the development of a new theoretical measure that can be extended and tested empirically (see Johns et al., 2016, for an empirical examination of contextual and semantic diversity). Theoretically, the justification for the importance of contextual diversity is provided by the rational analysis of memory framework (Anderson & Milson, 1989; Anderson & Schooler, 1991), which suggests that information in memory should be organized such that information that is more likely to be needed in the future is more available in memory. In terms of contextual diversity, this suggests that the best way to determine the likelihood of a word occurring in a future context is those that occurred in the greatest number of previous contexts and thus should be the words that are most available in the mental lexicon (see Jones et al., 2017, for an in-depth discussion of the relation between rational analysis and contextual diversity). Corpus-based modelling allows for a determination of the types of contexts that words occur, and thus is capable of constructing better estimates of a word's likely need. This research area demonstrates that by combining newly available resources (large-scale text corpora) with classic theories from cognitive psychology (the rational analysis of memory), new and better theories of cognition can be constructed.

Currently, the best contextual diversity measures are those derived from those utilizing the semantic distinctiveness model (SDM), first introduced by Jones et al. (2012), and has received multiple implementations since (Johns, 2021a; Johns et al., 2016; Johns, Jamieson, & Jones, 2020). The SDM modifies a contextual diversity measure by weighting each occurrence of a word in a context depending on how unique a contextual occurrence for a word is, with more unique contextual occurrences leading to a stronger encoding strength for that word. Uniqueness is defined based on the semantic similarity of the representation of a word and the context that it occurred in. The implementation with the best fit is the model of Johns (2021a; see Johns & Jones, 2022, for a further examination of this model), which used a communicative definition (a single individual's history of commenting on an internet discussion forum) of a linguistic context. This model was found to provide an 8.5%–22.5% increase in variance accounted for over word frequency across a variety of mega-large sets of lexical organization data (Johns, 2022b; Johns & Jones, 2022), while minimizing or eliminating the contribution of word frequency, demonstrating that this model is substantially more powerful in accounting for this type of data than word frequency. Indeed, the explanatory power of the SDM model of Johns (2021a) has been shown to generalize across behavioural data types, including lexical semantics (Johns, 2021b), recognition memory (Johns, 2022b), idiomatic processing (Senaldi et al., 2022), semantic decision (Antal et al., 2022), and cognitive aging (Johns et al., in press).

The ability of the SDM to construct a modified contextual diversity measure comes from the mechanisms of distributional models of semantics (see Günther et al., 2019; Kumar, 2021, for recent reviews). The standard information source that distributional models capitalize upon is that words that co-occur together within linguistic contexts tend to be semantically related, and that principle extends to first, second, and higher order contextual co-occurrence.

By learning word co-occurrence relationships across a very large corpus of text, distributional models construct accurate representations of the meaning of words and do so mechanistically via association instead of a rational symbolic process.

Distributional models are the greatest success story of scaled approaches to cognitive modelling, with the LSA model of Landauer and Dumais (1997) being the classic model. Additionally, LSA is an example of how technological innovations from computer science can inform the development of cognitive theory, as the underlying mechanisms that drive the model were developed to accomplish information retrieval from databases and were entitled latent semantic indexing (Deerwester et al., 1990). LSA constructs semantic representations for words by constructing a Word × Document matrix from a text corpus and decomposing it to construct a smaller semantic feature space through a technique entitled singular value decomposition. By taking the similarity between the inferred features of words, Landauer and Dumais (1997) demonstrated that LSA could account for a variety of different lexical semantic behaviours.

Since Landauer and Dumais (1997) was published, a variety of different distributional models of semantic cognition have been proposed, such as probabilistic inference methods (Griffiths et al., 2007), vector-accumulation/noise-cancellation methods (e.g., Jones & Mewhort, 2007), predictive neural networks (Mikolov et al., 2013), count-based methods (e.g., Bullinaria & Levy, 2007, 2012; Johns, Mewhort, et al., 2019; Lund & Burgess, 1996), and retrieval-based mechanisms (Jamieson et al., 2018; Johns & Jones, 2015; Kwantes, 2005). The success of this model type demonstrates the systematic connection between contextual word usage and the word meanings that people have acquired. That is, in contrast to the constrained training materials used in the approach of models employing the simplification assumption, distributional models learn the meaning of words from large-scale, noisy, and naturalistic information.

Cognitively Inspired Distributional Models

While the development of LSA came from research being done on information retrieval (similarly, the neural embedding model word2vec was developed by Google to accomplish mainly applied tasks before being adopted by cognitive science researchers; Mikolov et al., 2013), other distributional models have been developed from previously proposed cognitive models. A notable example of this is the BEAGLE model of Jones and Mewhort (2007). BEAGLE is a vector accumulation model, where a semantic representation is formed continuously based on how a word was used in a sentence. The learning mechanisms that the model utilized were based on the theory of distributed associative memory (TODAM) model of Murdock (1982), which is a computational model of episodic memory with an emphasis on understanding the storage of associative information.

In TODAM, two types of information are used to construct an episodic memory trace: item and associative information. Item information signals the occurrence of an item on a memory list, whereas associative information signals the occurrence of two words occurring together in a pair. To encode associative information, a technique known as convolution is used (with different implementations using different techniques for convolution). The current standard technique is entitled circular convolution, which is a function that takes in two vectors and constructs a new, unique vector that represents the association between the two. For example, if the model were simulating a paired-associate learning task (a task where participants have to remember pairs of words), and one pair was "*dog–cat*," in TODAM the vectors for dog and cat would be added into an episodic memory trace (to signal that those words occurred in the last), as well as the convolution of *dog* and *cat* (represented by $dog \circledast cat$ in Jones & Mewhort, 2007) to signal that those words are a pair. In this manner, an episodic trace is constructed such that both item and associative information are contained.

The BEAGLE model uses the same mechanisms proposed by TODAM to learn multiple aspects of word meanings, namely item (or context) and order information. However, instead of forming episodic traces of this information, a word's representation in the lexicon is constructed from processing sentences across a corpus (i.e., the episodic experiences that a word was used in).

Unlike other distributional models, such as LSA, BEAGLE operates at the sentence level of linguistic context. There are two types of vectors that are used by the model: environmental and memory vectors. Environmental vectors are static, meaning that they do not change across learning, and are used to mark a word's occurrence in a sentence. Each word in the model's lexicon has a separate environmental vector, and they are formed by randomly sampling from a Gaussian distribution. Memory vectors are dynamic, as they change across learning, and following the proposals of TODAM, there are both item and order memory vectors.

To build the memory vectors across learning, the model "reads" a sentence one word at a time and updates its memory representation in response to the word's occurrence. Item information is updated by summing the environmental vectors of the other words in the sentence (a window size parameter can be used to determine how many words are included in an update) into the word's item representation. A word's order representation is updated by adding all of the *n*-grams that surround a word in a sentence (up to a set size) using noncommutative circular convolution (Plate, 1995), and these *n*-gram vectors are added into the word's order representation. This is done for each word in a sentence and each sentence in a corpus. Thus, the word's item representation encodes pure co-occurrence, whereas the order representation encodes simplified syntactic information about how a word is used in relation to other words. It is common to sum these represents into a single composite representation. The ability to naturally acquire both co-occurrence and simple syntactic information is a unique capability of BEAGLE among distributional models. There is an alternative implementation of the model that utilizes sparse representations (Recchia et al., 2015; see Kanerva, 2009, for a discussion of the biological plausibility of this approach), which reduces the computational requirements of the model.

BEAGLE has proven to be able to successfully account for a variety of different empirical findings, including memory search (Hills et al., 2012), verbal fluency performance (Taler et al., 2020; Taler & Johns, 2022), semantic priming (Hare et al., 2009; Jones et al., 2006), changes in memory performance in clinical populations (Johns et al., 2018), individual and demographic differences in language processing (Aujla, 2021; Johns & Jamieson, 2018, 2019), and episodic memory effects (Mewhort et al., 2018; Osth et al., 2020), among others. When BEAGLE is combined with training methodologies from advanced machine learning algorithms, the performance of the model is similar to the top models in natural language processing (Johns, Mewhort, et al., 2019). Additionally, the simplicity of the model makes it possible to be optimized with advanced methodology targeting the training materials that are provided to the model, further increasing the performance of the model in comparison to other models (Johns, Jones, et al., 2019).

The success of BEAGLE demonstrates that the development of large-scale cognitive models does not have to come from research done in computer science, such as was done with LSA and word2-vec. Instead, standard cognitive mechanisms can be used to scale a model up to process similar amounts of information that an adult human being may have experienced. This suggests that classic mechanisms that have been developed over decades of theoretical research in the cognitive sciences are useful not only in the continued development of advanced cognitive theory but can also aid in the development of new technologies in machine learning and artificial intelligence.

Related to BEAGLE scaling up TODAM, the retrieval operations underlying MINERVA 2 have been used to develop a different conceptualization of semantic memory (Jones, 2019), through a combination of the processing mechanisms employed by MINERVA 2 and the representation assumptions of BEAGLE. In instance-based distributional models (e.g., Jamieson et al., 2018; Kwantes, 2005; see also Ambridge, 2020, for a general discussion of these issues). In this model type, there is no centralized representation of a word's meaning stored in memory, such as what is done in the majority of other distributional model types. Instead, each piece of language that a model encounters (whether it be at the sentence or document level) is stored as a single instance in memory. To retrieve the meaning of a word, the word is treated as a memory probe, and all of the traces that a word occurred in can be retrieved through a similarity function. The major advantage of this model type is that multiple words can be used as cues, enabling for combined meanings to be retrieved, and allowing for parsimonious explanations of complex linguistic phenomena such as polysemy.

Recent models by Johns and Jones (2015) and Johns, Jamieson, et al. (2020) have combined the representational framework of BEAGLE with the cued retrieval operations of MINERVA 2 in order to account for both sentence comprehension and production (see also Jamieson & Hauri, 2012; Jamieson & Mewhort, 2011, for an initial combination of these models in order to examine artificial grammar learning). In these models, the storage mechanisms from BEAGLE are used to construct realistic representations of sentence structures, and the retrieval operations from MINERVA 2 are used to construct expected syntactic structure that surrounds the usage of words, but with no grammatical information actually being integrated into the model's representation. Through this combination of processing and representation mechanisms, a wide variety of effects could be accounted for across both language comprehension and production. The combination of processing and representational models employed by these models signals a promising pathway towards more powerful models of cognition.

One specific challenge that has faced distributional models is the lack of sensory and motor information in their representations, a key limitation given the importance of this information type in theories of language and cognition (Barsalou, 1999, 2008). This is commonly referred to as the grounding problem. In recognition of this issue in distribution modelling (e.g., Riordan & Jones, 2011; see Wingfield & Connell, 2022, for a recent review), there has been a sustained effort to develop distributional models that integrate multisensory information into the word meaning representations constructed with distributional models (e.g., Andrews et al., 2009; Banks et al., 2021; Bruni et al., 2014; De Deyne et al., 2021; Johns & Jones, 2012; Lazaridou et al., 2017). Although the models all differ in their proposals as to the mechanisms by which distributional representations are grounded, all approaches integrate multimodel perceptual information into word meaning representations, typically through the utilization of various feature norms (e.g., Lynott et al., 2020; McRae et al., 2005). Given the substantial task faced by researchers attempting to ground distributional models, this will remain an important area of research for computational modelers and the cognitive realism of this model type. Relatedly, Johns (2021a, 2021b, 2022b) and Johns and Jones (2022) have demonstrated how social and communicative information can be integrated into models of lexical organization and lexical semantics, in order to ground these models in the social world, another important theoretical growth point for this type of model.

Integrating Representation and Processing

One of the major promises of distributional models is that they provide a representation upon which processing mechanisms can operate. As discussed, the vast majority of previous cognitive model efforts, with some notable exceptions (e.g., Nosofsky, 1986) have focused on processing assumptions. The outcome of training a distributional model is a vector-based semantic representation for each word in the lexicon. The most obvious use of these representations is to use them to directly test how the similarity structures of the learned representations map onto behavioural tasks examining lexical semantic properties of words (e.g., word similarity tasks). However, they can also be used to underlie a processing model, enabling realistic semantic representations to be integrated into a model's operations.

The most natural area of research where this integration is useful in the study of episodic memory. As discussed, computational modelling has a rich history in theory development in episodic memory. However, much of the focus has been placed on processing mechanisms rather than representational aspects of memory (mostly due to limitations in place on developing realistic representations of word meanings). By integrating a realistic representation of word meanings into a processing model, the similarity structure contained in semantics can be integrated into the model and be used to examine item-level effects in episodic memory. The most obvious application of this model is to empirical tasks that have a large influence on semantics.

A good example of this type of task is in the Deese/Roediger–McDermott (DRM) false memory paradigm (Deese, 1959; Roediger & McDermott, 1995), where participants are presented with lists of words that are all related to a single critical word (e.g., *hospital*, *nurse*, *medicine*, for the critical word *doctor*). The critical word is not presented to participants, but on tests of their memory, participants endorse critical words as being remembered at equal rates to studied words on both recognition and recall rates. Johns et al. (2012) constructed a model of both true and false recognition with a representation derived from a distributional model and a process model inspired by neural synchronization (Singer, 1999) and the mechanisms proposed by fuzzy trace theory (Brainerd & Reyna, 2002). This model, entitled the recognition through semantic synchronization model, can account for a variety of false memory effects, including item-level variability in the amount of false recognition that different DRM lists elicit (Gallo & Roediger, 2002; Roediger & McDermott, 1995), effects of associative and thematic strength (Cann et al., 2011), and developmental reversals in false recognition (Brainerd et al., 2002), among others. Additionally, Johns et al. (2021) recently expanded the model to also account for recollection-based data (see also Reid & Jamieson, 2022, for an application of the method to false recognition of unstudied but related lures).

A related area where integrated process and representation models have enabled theoretical development is verbal fluency and memory search. In a verbal fluency task, participants are asked to produce as many words as possible within a specific category in 1 min and is a commonly used tasked in neuropsychology examining language and memory impairments (Taler & Phillips, 2008). In examinations of semantic memory performance category fluency is used where participants are asked to produce as many words as possible from a given semantic category. Typically, performance in this task is measured by counting the number of category exemplars a participant produced, or with hand-coded measures of searching patterns within and across clusters within semantic space (Troyer et al., 1997).

The first model to explain category fluency using word representations derived from a distributional model was the animal foraging model of Hills et al. (2012; see also Hills et al., 2015), which used word representations derived from BEAGLE. The results of Hills et al. (2012) demonstrated that the memory search mechanisms that humans use to search through semantic memory seem to be similar to optimal foraging algorithms that animals use to find food in their physical environments (see Avery & Jones, 2018, for a follow up to this research, and Lundin et al., 2022, for an examination into the neurological underpinnings of optimal foraging in memory search). Similar research has been used to examine language switching in bilinguals (Taler et al., 2013), changes that occur during the development of a memory disorder (Johns et al., 2018), bilingual verbal fluency (Taler & Johns, 2022), and the dynamics of language production across the aging spectrum (Taler et al., 2020).

An additional area of research where representations derived from distributional models have been successfully employed is in judgement and decision-making (see Bhatia et al., 2019 for a recent review). For example, Bhatia (2017) demonstrated that the similarity structure of words that distributional models construct accurate predictions about the judgements that people make across a variety of tasks. These results have been generalized to other types of judgement tasks, such as numerical estimation (Zou & Bhatia, 2021), preferential choice tasks (Bhatia, 2019), and choice tendencies in a naturalistic data set (Bhatia & Walasek, 2019).

Combined, this work demonstrates the power and promise of using experientially scaled representations to drive cognitive models. Distributional models allow for representations to be constructed from realistic levels of experience. By using these representations in cognitive models, it allows for additional variance to be accounted for across behavioural data sets by allowing for item-level variance to be accounted for. Additionally, it accounts for lifespan effects by manipulating the amount and type of experience that a model can integrate into its representation (see Johns et al., 2012, 2019; Qiu & Johns, 2020; Taler et al., 2013, for examples).

This discussion of cognitively inspired distributional models and integrated process-representational models demonstrates the power of scaled approaches to computational cognitive modelling. However, the development of these models has not been completely independent of standard methodologies employing the simplification assumption. For example, the SDM model of Jones et al. (2012) was developed and validated in response to an artificial language experiment (and later with a mixed artificial/natural language experiment; Johns et al., 2016). Jamieson et al. (2018) and Crump et al. (2020) evaluated the unique capabilities of a retrieval-based distributional modelling using simple artificial languages. Asr and Jones (2017) used artificial languages to contrast and compare different distributional modelling approaches. Similarly, Mannering and Jones (2021) used artificial languages to demonstrate issues of catastrophic interference in neural embedding distributional models. Artificial languages were used in these examples due to the simplification assumption provide fine-grained capabilities to isolate how a model is operating (McClelland, 2009; Shiffrin, 2010) while limiting representational complexity. This can be difficult to accomplish when using large natural corpora, as it is difficult to determine why a model is behaving as it is when learning from large and noisy data. Thus, scaled and simplified approaches to cognitive modelling are not opposite approaches, as both offer their advantages and disadvantages, and researchers should be aware of the strengths and limitations of both approaches.

Methodological Issues

Using distributional models as an underlying representation in a cognitive model presents unique challenges and opportunities in contrast to classical processing-focused approaches. One of the major challenges facing the development of new scaled models is the training materials that are used, as these models are experientially dependent. Distributional models face a classic problem in computational systems, best summarized with the expression "garbage in, garbage out." The quality of the representations that distributional models can construct is directly dependent on the informational content of the text corpora that a model is given.

For applied research developing intelligent systems, the answer to this problem is relatively simple—use the one that seems to provide the best solution within a given problem domain. Typically, this manifests in using the greatest amount of training materials possible, such as what is done with transformer networks like General Pre-Trained Transformer (GPT-3; Brown et al., 2020), which is trained on a very large subsection of the entire internet. However, psychologists looking to develop cognitive theory from distributional models face a more complex task in deriving correct training materials. Specifically, the promise of distributional modelling in the cognitive sciences is that they allow for lifespan-level experiences with language to be built into a cognitive model. The types of experiences that individuals have with language are varied and diverse, which are not necessarily taken into account by the typical training materials that distributional models employ (e.g., collections of fiction books or Wikipedia articles are likely not reflective of the actual linguistic experience of individuals).

To gain an initial understanding of the impact of training materials on distributional model performance, Johns and Jamieson (2018) examined a large set of fiction books organized by author and genre and found that an individual author's use of language was very distinct and was much greater than the impact of genre. This result was replicated by Johns, Dye, et al. (2020) with a much larger set of books. This finding suggests that individuals vary greatly in their usage of language and begs the question as to what leads to this variability.

Johns, Jones, et al. (2019) took a more quantitative approach to this question and devised an algorithm that optimized a distributional model's performance by selecting pieces of language that allowed the model to have the best fit to a set of data. The method is entitled EO. The motivation for the development of EO was to determine if the lexical experience that different individuals have is embedded in their lexical behaviours. It was found that using EO provided multiple models the ability to achieve benchmark fits to various types of data across lexical semantics, lexical organization, sentence processing, and episodic recognition. Importantly, it was found that EO could infer demographic information about the participants from whom the data were collected. Specifically, when fitting to lexical decision data collected from younger and older adults, it was found that the method preferred young adult novels when fitting to younger adult data and advanced fiction novels when fitting to older adult data. This was further validated by Johns and Jamieson (2019) who found the method preferred time and place appropriate language sources when fitting to data collected from different times and places.

From a high-level perspective, these results indicate that since distributional models are experientially dependent, they should be trained with materials that are consistent with the types of linguistic experiences of the participants whose data are being investigated. Additionally, this suggests that when using or comparing distributional model performance, close attention should be paid to the types of training materials used to optimize a model, as a model may be incorrectly rejected if it was not provided with appropriate samples of text. An important goal of distributional models should be the continued collection of different sources of text that correspond to people's experience with the language environment.

A related issue to the type of language that is contained in a training corpus is that of corpus length. Although it is common practice to train a distributional model with a maximal amount of linguistic materials (e.g., all of Wikipedia), this is not reflective of the amount of linguistic experience that an average person has received. Although it is difficult to estimate the amount of language one experiences across their lifespan, studies (Brysbaert et al., 2016; Mehl et al., 2007) estimate that an average student experiences about 12 million word tokens a year. For an average 60-year-old, this works out to about 700 million word tokens experienced across their lifespan. Thus, while humans receive a very substantial amount of linguistic, it likely does not reach to the billions of words that are commonly used in distributional model training. In Johns, Mewhort, et al. (2019), when using EO to optimize a model, the method only selected a limited number of language sections (although the number selected is dependent on the number of words used in a behavioural data set), suggesting that language quality is more important than amount of language in improving distributional model performance. However, when training distributional models to explain human cognition, attention should be paid to corpus size, since if the model requires more training materials than is plausible for a human to receive that model is likely not a good candidate as a plausible model of cognition. However, corpus-based analyses are not just used in the development of cognitive models but can also be used to determine the overall structure of the language environment (e.g., Johns, 2021a). In these cases, maximizing the amount of materials may provide more insight into the overall shape of the language that people could experience.

An additional issue that needs to be considered in scaled cognitive modelling is model complexity. The complexity of a cognitive model is typically quantified as some transformation of the parameter space of the model (e.g., Shiffrin et al., 2008), stemming from classic model testing procedures such as Bayesian information criterion (Schwarz, 1978) or Akaike information criterion (Akaike, 1973), both of which penalize model fit as a function of the number of available free parameters. As with other areas of cognitive modelling, distributional models differ in their parameter space. For example, neural embedding models (e.g., word2vec; Mikolov et al., 2013) have a much larger parameter than most distributional models (Johns, Mewhort, et al., 2019). Other approaches, such as pointwise mutual information (Bullinaria & Levy, 2007, 2012) have an extremely small parameter space. Although little attention has been given to the relative costs and benefits of including free parameters across the different distributional models, it has been repeatedly found that by taking effort to equate free parameters in different models with similar training assumptions, different distributional models tend to perform within the same range of success (Johns, Mewhort, et al., 2019; Levy et al., 2015).

Another source of complexity that should be considered in distributional modelling is the computational complexity of different model types (Recchia & Jones, 2009). Computational complexity comes from theoretical computer science and is based upon the number of steps an algorithm requires, typically in the worse case, given an input of a certain size. Different distributional model types have different computational complexity levels—for example, LSA is considerably more computationally complex than pointwise mutual information or BEAGLE. This has been proposed to be a general consideration in cognitive science (Beal & Roberts, 2009) and one that is often ignored. Just as complexity of the model parameter space has been considered an insight into the underlying flexibility that a model has in fitting to data, computational complexity may have an equivalent role to play in increasing model power. However, this is an understudied problem that should be a question for future research (see Jones & Dzhafarov, 2014, for an example of similar issues within a different domain).

The issue of computational complexity is related to a general issue in scaled approaches to cognitive modelling—that of being able to derive sound theoretical understanding and cognitive principles from a model. This is an issue discussed by McClelland (2009), who provided a parable of mapmakers. This parable, initially proposed by Borges (1998), describes how if one is obsessed with making the most realistic maps, the end result is a map that rivals the size of the real world, a not overly useful endeavor. Thus, a practical map is one that is abstracted from the geographic world enough to provide a reasonable guide for navigation. Similar issues arise in scaled cognitive modelling when it becomes difficult to determine the underlying reasons for a model's success due to issues of underlying complexity of the models.

A good example of this issue is provided in the neural embedding model word2vec (Mikolov et al., 2013). The original article describing this model is sparse on detail and evaluation, which prompted the need for other researchers to examine and understand the nature of the model's operations (e.g., Levy & Goldberg, 2014). One specific example of understanding word2vec's operations is the use of negative sampling, where the network is updated to be unpredictive of unrelated words, accomplished through the generation of words based on precomputed word frequency values. Goldberg and Levy (2014) sought to understand the impact of this procedure through mathematical analysis, with their ultimate conclusion reading, "Why does this produce good word representations? Good question. We don't really know." (p. 5). The impact of negative sampling was not determined until Johns, Mewhort, et al. (2019) used a simpler distributional framework to understand the impact of negative sampling, where it was determined the success of the procedure was due to negative sampling integrating baseline levels of co-occurrence into a word's representation, allowing for unique co-occurrence values to be highlighted. This finding differed from the common assumption that negative sampling was being used to hone a prediction method.

This issue becomes even more pertinent when discussing current state-of-the-art deep learning models emerging in artificial intelligence and machine learning research, such as the GPT network (Floridi & Chiriatti, 2020). This model, while extremely impressive in performance, has a parameter space in the billions and uses many different operations to hone and optimize its performance, and is trained on the text of almost the entire internet (i.e., an amount of training materials that is orders of magnitude larger than a typical human receives). Thus, understanding the reasons for the model's behaviour is nearly impossible compared to standard approaches in cognitive modelling. Indeed, researchers are now collecting human data to try to understand the connection between human cognition and GPT and to elucidate how and why the model is successful (e.g., Goldstein et al., 2022). If one's model is so complex that it requires data collection to understand its performance, it is likely not informative for theory generation in the cognitive sciences, as it has simply shifted the goalposts from understanding how human's operate to understanding how a machine that behaves somewhat like a human operates (see Jones et al., 2015, Johns, Jamieson, & Jones, 2020, for a related discussion of these issues).

From a psychological perspective, the scaling up of classic cognitive models such as TODAM and MINERVA 2 eliminates these issues, as these are well-established models of cognition that have been used, studied, and evaluated by cognitive psychologists for decades. Although these models are not as optimized and powerful as the larger machine learning models, they provide theoretical clarity into the contribution of different learning and retrieval mechanisms in explaining human behaviour. Thus, psychologists should not throw away their models, developed to explain well-defined laboratory tasks, but instead should determine how these models can be integrated with scalable learning mechanisms to evaluate their performance when presented with large and noisy amounts of training materials.

General Discussion

The goal of this article was to review common practices within cognitive modelling and to determine the role that machine learning and big data technologies could play in developing cognitive theory. Specifically, this article has focused on the role that representation, and in turn memory content, plays in explaining human cognition. In the history of computational cognitive modelling, there have been multiple strategies used to integrate content into the computational underpinnings of a model. One approach has been to generate artificial stimuli that seek to capture some aspects of complex training materials, such as natural language, in order to focus the modelling exercise on the processing mechanisms that control and produce behaviour (e.g., Elman, 1990; Reber, 1976; Rogers & McClelland, 2004). This approach to cognitive modelling is best exemplified by the simplification assumption (McClelland, 2009; Shiffrin, 2010), which seeks to explain away complexity in one domain (e.g., environmental occurrence patterns) to focus on certain aspects of a model's framework (e.g., learning or processing mechanisms) to gain a more complete understanding of a model's success. The simplification assumption has played a central role in the history of cognitive modelling in generating sophisticated models of human cognition.

Scaled cognitive modelling offers an alternative to the use of the simplification assumption, as they provide an ability to generate realistic memory representations that contain content that is reflective of the experience that an individual may receive during their lifetime. Similar to Simon's (1969) parable of the ant, scalable modelling aims to evaluate a cognitive mechanism in the context of the rich environmental regularities that it evolved to thrive in. The integration of realistic representation types with plausible processing mechanisms provides a promising pathway for the development of more powerful, and increasingly plausible, models of cognition. As detailed, scaled cognitive models have proven to be successful at accounting for a wide variety of different behavioural data types, such as morphological processing (e.g., Marelli & Baroni, 2015; Marelli et al., 2017), lexical organization (e.g., Hoffman et al., 2013; Johns, 2021a, 2022b; Jones et al., 2012), episodic memory (e.g., Johns et al., 2012, 2021; Mewhort et al., 2018; Osth et al., 2020), verbal fluency (e.g., Hills et al., 2012; Taler et al., 2013), sentence processing (e.g., Johns & Jones, 2015), language production (Johns, Jamieson, et al., 2020), and decision (e.g., Bhatia, 2017), among others.

This multitude of findings is coherent with classic goals in the cognitive sciences (Estes, 1955; Simon, 1969), which propose that cognitive models should include environmental structure to simplify their operations and increase the model's adaptability. New approaches to generating cognitive theory using scaled cognitive models offer an increased ability to integrate this information into a model's performance, something that has not been available to previous generations of researchers due to a multitude of technological limitations. However, given the ability of distributional models to generate realistic representations of word-based stimuli, it provides an ability to understand the interaction between process and representation in understanding cognition, an issue of central importance in generating cognitive theory (Castro & Siew, 2020; Jamieson et al., 2022).

A central goal of the psychological and cognitive sciences is to explain human behaviour across the lifespan (Wulff et al., 2019). As people age, lexical experience is accumulated within memory which has a causal impact on behaviour at different ages (Ramscar et al., 2014, 2017; see Ramscar, 2022, for a review). Most traditional approaches to cognitive modelling ignore this issue (although see Ramscar et al., 2014, 2017, for an alternative perspective). Distributional and corpus-based models of cognition offer an ability to include accumulated experience into a variety of models explaining behaviour at different time points, for example, in paired-associate learning (Qiu & Johns, 2020) and verbal fluency (Taler et al., 2020) performance across the aging spectrum. Basing a cognitive model's operations on a realistic representation type allows it to be tested on a greater variety of data, such as developmental and aging data.

A related overarching question is when a researcher should consider using a scaled representation to integrate with a processing model, as there are tasks where using these representations would not lead to significant increases in model performance. For example, if a task focused on orthographic or phonological processing, the integration of a distributional representation may not provide much more power to a model than traditional techniques offer. It is difficult to predict a priori the impact of past experience on task performance, which has led some researchers (e.g., Nelson & Shiffrin, 2013) to use previously unseen stimuli in their experimental designs. However, the integration of a scaled representation into a process model's behaviour provides an interesting pathway for model testing procedures, as it provides a method for researchers to definitively determine whether a task is influenced by the meaning of words, which could provide further insights into the cognitive mechanisms that are at play for different behavioural data types.

Historically, in cognitive modelling, more attention has been paid to process overrepresentation. This has been mainly due to technological limitations in generating realistic representation types. New corpus-based models allow for the generation of realistic representations that contain semantic content to be built into a model's processing, which allows for behaviour to be accounted for with an interaction between process and representation. However, as with any new technological improvement in the cognitive sciences, the development of representational models needs to be crouched within standard cognitive theory in order to ensure a productive and cumulative science.

Résumé

L'objectif classique de la modélisation cognitive est l'intégration du processus et de la représentation pour développer des théories complètes de la cognition humaine (Estes, 1955). Cet objectif est encore mieux exprimé par le travail précurseur de Simon (1969) qui s'est servi de la parabole de la fourmi pour illustrer l'importance de comprendre l'environnement dans lequel une personne est intégrée au moment d'élaborer une théorie de la cognition. Cependant, les hypothèses habituelles pour expliquer le rôle de la représentation dans les modèles cognitifs computationnels ne représentent pas exactement le contenu de la mémoire (Johns et Jones, 2010). Le développement récent des approches de la cognition fondées sur l'apprentissage automatique et les mégadonnées, désignées ici par le terme de modélisation cognitive à l'échelle, offre une solution potentielle à l'intégration du processus et de la représentation. Cet article examine les pratiques et les hypothèses standard qui interviennent dans la modélisation cognitive, et comment les nouvelles approches axées sur les mégadonnées et l'apprentissage automatique modifient ces pratiques, de même que les directions que devraient prendre les recherches à venir. L'article a pour but d'ancrer les approches axées sur les mégadonnées et l'apprentissage automatique dans les sciences cognitives et dans les principes théoriques cognitifs classiques, afin de dégager une voie constructive vers l'intégration de la théorie cognitive à la méthodologie informatique avancée.

Mots-clés : modélisation cognitive, apprentissage automatique, mégadonnées, sémantique lexicale, modélisation distributionnelle

References

Adelman, J. S., & Brown, G. D. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review*, *115*(1), 214–227. https://doi.org/10.1037/0033-295X.115.1.214

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*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the second international symposium on information theory* (pp. 267–281). Akademiai Kiado.

Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, *40*(5–6), 509–559. https://doi.org/10.1177/0142723719869731

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*(4), 703–719. https://doi.org/10.1037/0033-295X.96.4.703

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408. https://doi.org/10.1111/j.1467-9280.1991.tb00174.x

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. https://doi.org/10.1037/a0016261

Antal, C., de Almeida, R. G., & Johns, B. T. (2022). Which semantic properties of a feature affect access to an object concept? In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual conference of the cognitive science society* (pp. 1–7). Cognitive Science Society.

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language*, *39*(3), 371–391. https://doi.org/10.1006/jmla.1998.2581

Asr, F. T., & Jones, M. N. (2017). *An artificial language evaluation of distributional semantic models* [Conference session]. Proceedings of the ACL Conference on Natural Language Learning (CoNLL), Vancouver, Canada.

Asr, F. T., Willits, J. A., & Jones, M. N. (2016). *Comparing predictive and co-occurrence based models of lexical semantics trained on child-directed speech* [Paper presentation]. Proceedings of the 37th Meeting of the Cognitive Science Society, Austin, Texas, United States.

Aujla, H. (2021). Language experience predicts semantic priming of lexical decision. *Canadian Journal of Experimental Psychology*, *75*(3), 235–244. https://doi.org/10.1037/cep0000255

Avery, J. E., & Jones, M. N. (2018). *Comparing models of semantic fluency: Do humans forage optimally or walk randomly?* [Conference session]. Proceedings of the 40th Annual Meeting of the 40th Annual Meeting of the Cognitive Science Society.

Banks, B., Wingfield, C., & Connell, L. (2021). Linguistic distributional knowledge and sensorimotor grounding both contribute to semantic category production. *Cognitive Science*, *45*(10), Article e13055. https://doi.org/10.1111/cogs.13055

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Beal, J., & Roberts, J. (2009). *Enhancing methodological rigor for computational cognitive science: Complexity analysis* [Paper presentation]. Proceedings of the Annual Meeting of the Cognitive Science Society, Austin, Texas, United States.

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, *124*(1), 1–20. https://doi.org/10.1037/rev0000047

Bhatia, S. (2019). Semantic processes in preferential decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 627–640. https://doi.org/10.1037/xlm0000618

Bhatia, S., Richie, R., & Zou, W. (2019). Distributed semantic representations for modeling human judgment. *Current Opinion in Behavioral Sciences*, *29*, 31–36. https://doi.org/10.1016/j.cobeha.2019.01.020

Bhatia, S., & Walasek, L. (2019). Association and response accuracy in the wild. *Memory & Cognition*, *47*(2), 292–298. https://doi.org/10.3758/s13421-018-0869-6

Borges, J. L. (1998). On the exactitude of science (A. Hurley, Trans.), *Collected fictions* (p. 325). Penguin.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind: Discoveries in Cognitive Science*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*(5), 164–169. https://doi.org/10.1111/1467-8721.00192

Brainerd, C. J., Reyna, V. F., & Forrest, T. J. (2002). Are young children susceptible to the false-memory illusion? *Child Development*, *73*(5), 1363–1377. https://doi.org/10.1111/1467-8624.00477

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, *74*(1), 1–15. https://doi.org/10.1037/h0024206

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, *49*, 1–47. https://doi.org/10.1613/jair.4135

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., & New, B. (2009). Moving beyond Kucera 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*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*, Article 1116. https://doi.org/10.3389/fpsyg.2016.01116

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526. https://doi.org/10.3758/BF03193020

Bullinaria, J. A., & Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: Stop-lists, stemming, and SVD. *Behavior Research Methods*, *44*(3), 890–907. https://doi.org/10.3758/s13428-011-0183-8

Caldwell-Harris, C. L. (2021). Frequency effects in reading are powerful—But is contextual diversity the more important variable? *Language and Linguistics Compass*, *15*(12), Article e12444. https://doi.org/10.1111/lnc3.12444

Cann, D. R., McRae, K., & Katz, A. N. (2011). False recall in the Deese/Roediger–McDermott paradigm: The roles of gist and associative strength. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *64*(8), 1515–1542. https://doi.org/10.1080/17470218.2011.560272

Castro, N., & Siew, C. S. Q. (2020). Contributions of modern network science to the cognitive sciences: Revisiting research spirals of representation and process. *Proceedings—Royal Society. Mathematical, Physical and Engineering Sciences*, *476*(2238), Article 20190825. https://doi.org/10.1098/rspa.2019.0825

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Chomsky, N. (1991). Linguistics and cognitive science: Problems and mysteries. In A. Kasher (Ed.), *The chomskyan turn* (pp. 26–53). Blackwell.

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: Human Experimental Psychology*, *69*(6), 1049–1055. https://doi.org/10.1080/17470218.2015.1130068

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, *23*(3), 371–414. https://doi.org/10.1207/s15516709cog2303_4

Crump, M. J. C., Jamieson, R. K., Johns, B. T., & Jones, M. N. (2020) Controlling the retrieval of general vs specific semantic knowledge in the instance theory of semantic memory. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd annual conference of the cognitive science society* (pp. 3261–3267). Cognitive Science Society.

Davies, M. (2009). The 385+ million word corpus of contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*(2), 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

De Deyne, S., Navarro, D. J., Collell, G., & Perfors, A. (2021). Visual and affective multimodal models of word meaning in language and mind. *Cognitive Science*, *45*(1), Article e12922. https://doi.org/10.1111/cogs.12922

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452–478. https://doi.org/10.1037/0033-295X.108.2.452

Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Developmental Psychology*, *53*(8), 1560–1570. https://doi.org/10.1037/dev0000347

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Elman, J. L., & McRae, K. (2019). A model of event knowledge. *Psychological Review*, *126*(2), 252–291. https://doi.org/10.1037/rev0000133

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*(5), 369–377. https://doi.org/10.1037/h0046888

Feldman-Stewart, D., & Mewhort, D. J. K. (1994). Learning in small connectionist networks does not generalize to large networks. *Psychological Research*, *56*(2), 99–103. https://doi.org/10.1007/BF00419716

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

Gallo, D. A., & Roediger, H. L., III. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language*, *47*(3), 469–497. https://doi.org/10.1016/S0749-596X(02)00013-X

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8–20. https://doi.org/10.3758/BF03198438

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 5–16. https://doi.org/10.1037/0278-7393.16.1.5

Goldberg, Y., & Levy, O. (2014). *word2vec Explained: Deriving Mikolov et al.'s negative-sampling word-embedding method*. arXiv. https://doi.org/10.48550/arXiv.1402.3722

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., … Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, *25*(3), 369–380. https://doi.org/10.1038/s41593-022-01026-4

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186. https://doi.org/10.1016/S1364-6613(00)01467-4

Gould, S. J. (1996). *The mismeasure of man*. W.W. Norton.

Graves, A., Mohamed, A. R., & Hinton, G. (2013). *Speech recognition with deep recurrent neural networks* [Conference session]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, pp. 6645–6649.

Graves, A., & Schmidhuber, J. (2008). *Offline handwriting recognition with multidimensional recurrent neural networks* [Conference session]. Advances in Neural Information Processing Systems, Red Hook, New York, United States.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. https://doi.org/10.1037/0033-295X.114.2.211

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, *111*(2), 151–167. https://doi.org/10.1016/j.cognition.2009.01.009

Hebb, D. O. (1958). Alice in Wonderland or psychology among the biological sciences. In H. E. Hartow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior* (pp. 451–467). University of Wisconsin.

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*(4), 976–995. https://doi.org/10.1111/cogs.12392

Herrnstein, R. J., & Murray, C. (2010). *The bell curve: Intelligence and class structure in American life*. Simon and Schuster.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431–440. https://doi.org/10.1037/a0027373

Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in semantic fields: How we search through memory. *Topics in Cognitive Science*, *7*(3), 513–534. https://doi.org/10.1111/tops.12151

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551. https://doi.org/10.1037/0033-295X.95.4.528

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hollis, G., & Westbury, C. (2016). The principals of meaning: Extracting semantic dimensions from co-occurrence models of semantics. *Psychonomic Bulletin & Review*, *23*(6), 1744–1756. https://doi.org/10.3758/s13423-016-1053-2

Huebner, P. A., & Willits, J. A. (2018). Structured semantic knowledge can emerge automatically from predicting word sequences in child-directed speech. *Frontiers in Psychology*, *9*, Article 133. https://doi.org/10.3389/fpsyg.2018.00133

Jackendoff, R. S. (1992). *Semantic structures*. MIT Press.

Jackson, P. (1986). *Introduction to expert systems*. Addison-Wesley.

Jamieson, R. K., Avery, J. E., Johns, B. T., & Jones, M. N. (2018). An instance theory of semantic memory. *Computational Brain & Behavior*, *1*(2), 119–136. https://doi.org/10.1007/s42113-018-0008-2

Jamieson, R. K., Crump, M. J., & Hannah, S. D. (2012). An instance theory of associative learning. *Learning & Behavior*, *40*(1), 61–82. https://doi.org/10.3758/s13420-011-0046-2

Jamieson, R. K., & Hauri, B. R. (2012). An exemplar model of performance in the artificial grammar task: Holographic representation. *Canadian Journal of Experimental Psychology*, *66*(2), 98–105. https://doi.org/10.1037/a0027023

Jamieson, R. K., Johns, B. T., Vokey, J. R., & Jones, M. N. (2022). Instance theory as a domain-general framework for cognitive psychology. *Nature Reviews Psychology*, *1*(3), 174–183. https://doi.org/10.1038/s44159-022-00025-3

Jamieson, R. K., & Mewhort, D. J. K. (2009). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *62*(3), 550–575. https://doi.org/10.1080/17470210802055749

Jamieson, R. K., & Mewhort, D. J. K. (2010). Applying an exemplar model to the artificial-grammar task: String completion and performance on individual items. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *63*(5), 1014–1039. https://doi.org/10.1080/17470210903267417

Jamieson, R. K., & Mewhort, D. J. K. (2011). Grammaticality is inferred from global similarity: A reply to Kinder (2010). *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *64*(2), 209–216. https://doi.org/10.1080/17470218.2010.537932

Johns, B. T. (2019). Mining a crowdsourced dictionary to understand consistency and preference in word meanings. *Frontiers in Psychology*, *10*, Article 268. https://doi.org/10.3389/fpsyg.2019.00268

Johns, B. T. (2021a). Disentangling contextual diversity: Communicative need as a lexical organizer. *Psychological Review*, *128*(3), 525–557. https://doi.org/10.1037/rev0000265

Johns, B. T. (2021b). Distributional social semantics: Inferring word meanings from communication patterns. *Cognitive Psychology*, *131*, Article 101441. https://doi.org/10.1016/j.cogpsych.2021.101441

Johns, B. T. (2022a). *Determining the relativity of word meanings through the construction of individualized models of semantic memory*. PsyArXiv. https://doi.org/10.31234/osf.io/2kghn

Johns, B. T. (2022b). Accounting for item-level variance in recognition memory: Comparing word frequency and contextual diversity. *Memory & Cognition*, *50*(5), 1013–1032. https://doi.org/10.3758/s13421-021-01249-z

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, *23*(4), 1214–1220. https://doi.org/10.3758/s13423-015-0980-7

Johns, B. T., Dye, M., & Jones, M. N. (2020). Estimating the prevalence and diversity of words in written language. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *73*(6), 841–855. https://doi.org/10.1177/1747021819897560

Johns, B. T., & Jamieson, R. K. (2018). A large-scale analysis of variance in written language. *Cognitive Science*, *42*(4), 1360–1374. https://doi.org/10.1111/cogs.12583

Johns, B. T., & Jamieson, R. K. (2019). The influence of place and time on lexical behavior: A distributional analysis. *Behavior Research Methods*, *51*(6), 2438–2453. https://doi.org/10.3758/s13428-019-01289-z

Johns, B. T., Jamieson, R. K., Crump, M. J. C., Jones, M. N., & Mewhort, D. J. K. (2020). Production without rules: Using an instance memory model to exploit structure in natural language. *Journal of Memory and Language*, *115*, Article 104165. https://doi.org/10.1016/j.jml.2020.104165

Johns, B. T., Jamieson, R. K., & Jones, M. N. (2020). The continued importance of theory: Lessons from big data approaches to cognition. In S. E. Woo, R. Proctor, & L. Tay (Eds.), *Big data methods for psychological research: New horizons and challenges* (pp. 277–295). APA Books. https://doi.org/10.1037/0000193-013

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, *17*(5), 662–672. https://doi.org/10.3758/PBR.17.5.662

Johns, B. T., & Jones, M. N. (2012). Perceptual inference through global lexical similarity. *Topics in Cognitive Science*, *4*(1), 103–120. https://doi.org/10.1111/j.1756-8765.2011.01176.x

Johns, B. T., & Jones, M. N. (2015). Generating structure from experience: A retrieval-based model of language processing. *Canadian Journal of Experimental Psychology*, *69*(3), 233–251. https://doi.org/10.1037/cep0000053

Johns, B. T., & Jones, M. N. (2022). Content matters: Measures of contextual diversity must consider semantic content. *Journal of Memory and Language*, *123*, Article 104313. https://doi.org/10.1016/j.jml.2021.104313

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2012). A synchronization account of false recognition. *Cognitive Psychology*, *65*(4), 486–518. https://doi.org/10.1016/j.cogpsych.2012.07.002

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2019). Using experiential optimization to build lexical representations. *Psychonomic Bulletin & Review*, *26*(1), 103–126. https://doi.org/10.3758/s13423-018-1501-2

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2021). A continuous source reinstatement model of true and false recollection. *Canadian Journal of Experimental Psychology*, *75*(1), 1–18. https://doi.org/10.1037/cep0000237

Johns, B. T., Mewhort, D. J. K., & Jones, M. N. (2017). Small worlds and big data: Examining the simplification assumption in cognitive modeling. In M. N. Jones (Ed.), *Big data in cognitive science: From methods to insights* (pp. 227–245). Taylor & Francis.

Johns, B. T., Mewhort, D. J. K., & Jones, M. N. (2019). The role of negative information in distributional semantic learning. *Cognitive Science*, *43*(5), Article e12730. https://doi.org/10.1111/cogs.12730

Johns, B. T., Taler, V., & Jones, M. N. (in press). Contextual dynamics in lexical encoding across aging: A simulation study. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218221145685

Johns, B. T., Taler, V., Pisoni, D. B., Farlow, M. R., Hake, A. M., Kareken, D. A., Unverzagt, F. W., & Jones, M. N. (2018). Cognitive modeling as an interface between brain and behavior: Measuring the semantic decline in mild cognitive impairment. *Canadian Journal of Experimental Psychology*, *72*(2), 117–126. https://doi.org/10.1037/cep0000132

Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*(1), 1–32. https://doi.org/10.1037/a0034190

Jones, M. N. (2019). When does abstraction occur in semantic memory: Insights from distributional models. *Language, Cognition and Neuroscience*, *34*(10), 1338–1346. https://doi.org/10.1080/23273798.2018.1431679

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizational principle of the lexicon. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 239–283). Elsevier Academic Press.

Jones, M. N., Hills, T. T., & Todd, P. M. (2015). Hidden processes in structural representations: A reply to Abbott, Austerweil, and Griffiths (2015). *Psychological Review*, *122*(3), 570–574. https://doi.org/10.1037/a0039248

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, *66*(2), 115–124. https://doi.org/10.1037/a0026727

Jones, M. N., Kintsch, W., & Mewhort, D. J. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552. https://doi.org/10.1016/j.jml.2006.07.003

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. https://doi.org/10.1037/0033-295X.114.1.1

Kanerva, P. (2009). Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation*, *1*(2), 139–159. https://doi.org/10.1007/s12559-009-9009-8

Kucera, H., Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40–80. https://doi.org/10.3758/s13423-020-01792-x

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, *12*(4), 703–710. https://doi.org/10.3758/BF03196761

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Lazaridou, A., Marelli, M., & Baroni, M. (2017). Multimodal word meaning induction from minimal exposure to natural text. *Cognitive Science*, *41*(Suppl. 4), 677–705. https://doi.org/10.1111/cogs.12481

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Levy, O., & Goldberg, Y. (2014). *Neural word embedding as implicit matrix factorization* [Conference session]. Advances in Neural Information Processing Systems, Cambridge, Massachusetts, United States.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embedding. *Transactions of the Association for Computational Linguistics*, *3*, 211–225. https://doi.org/10.1162/tacl_a_00134

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lundin, N. B., Brown, J. W., Johns, B. T., Jones, M. N., Purcell, J. R., Hetrick, W. P., O'Donnell, B. F., & Todd, P. M. (2022). *Neural switch processes guide semantic and phonetic foraging in human memory*. PsyArXiv. https://doi.org/10.31234/osf.io/857he

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Mannering, W. M., & Jones, M. N. (2021). Catastrophic interference in predictive neural network models of distributional semantics. *Computational Brain & Behavior*, *4*(1), 18–33. https://doi.org/10.1007/s42113-020-00089-5

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, *122*(3), 485–515. https://doi.org/10.1037/a0039267

Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as abstract operation in semantic space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition*, *166*, 207–224. https://doi.org/10.1016/j.cognition.2017.05.026

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38. https://doi.org/10.1111/j.1756-8765.2008.01003.x

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*(3), 295–322. https://doi.org/10.1177/00238309010440030101

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BF03192726

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, *317*(5834), 82–82. https://doi.org/10.1126/science.1139940

Mewhort, D. J. (1990). Alice in wonderland, or psychology among the information sciences. *Psychological Research*, *52*(2–3), 158–162. https://doi.org/10.1007/BF00877524

Mewhort, D. J. K., Shabahang, K. D., & Franklin, D. R. J. (2018). Release from PI: An analysis and a model. *Psychonomic Bulletin & Review*, *25*(3), 932–950. https://doi.org/10.3758/s13423-017-1327-3

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*, 3111–3119.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609–626. https://doi.org/10.1037/0033-295X.89.6.609

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*(3), 721–756. https://doi.org/10.1037/0033-295X.111.3.721

Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, *120*(2), 356–394. https://doi.org/10.1037/a0032020

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*(2), 327–357. https://doi.org/10.1037/0033-295X.113.2.327

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2020). Search for the missing dimensions: Building a feature-space representation for a natural-science category domain. *Computational Brain & Behavior*, *3*(1), 13–33. https://doi.org/10.1007/s42113-019-00033-2

Osth, A. F., Shabahang, K. D., Mewhort, D. J., & Heathcote, A. (2020). Global semantic similarity effects in recognition memory: Insights from BEAGLE representations and the diffusion decision model. *Journal of Memory and Language*, *111*, Article 104071. https://doi.org/10.1016/j.jml.2019.104071

Pinker, S. (1994). *The language instinct: How the mind creates language*. William Morrow. https://doi.org/10.1037/e412952005-009

Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, *6*(3), 623–641. https://doi.org/10.1109/72.377968

Qiu, M., & Johns, B. T. (2020). Semantic diversity in paired-associate learning: Further evidence for the information accumulation perspective of cognitive aging. *Psychonomic Bulletin & Review*, *27*(1), 114–121. https://doi.org/10.3758/s13423-019-01691-w

Quine, W. V. O. (1960). *Word and object*. MIT Press.

Ramscar, M. (2022). *Psycholinguistics and aging*. Oxford Research Encyclopedia of Linguistics. https://doi.org/10.1093/acrefore/9780199384655.013.374

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, *6*(1), 5–42. https://doi.org/10.1111/tops.12078

Ramscar, M., Sun, C. C., Hendrix, P., & Baayen, H. (2017). The mismeasurement of mind: Life-span changes in paired-associate learning scores reflect the "cost" of learning, not cognitive decline. *Psychological Science*, *28*(8), 1171–1179. https://doi.org/10.1177/0956797617706393

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *2*, 88–94. https://doi.org/10.1037/0278-7393.2.1.88

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*, *41*(3), 647–656. https://doi.org/10.3758/BRM.41.3.647

Recchia, G. L., Sahlgren, M., Kanerva, P., & Jones, M. N. (2015). Encoding sequential information in vector space models of semantics: Comparing holographic reduced representation and random permutation. *Computational Intelligence and Neuroscience*, *2015*, 58–58. https://doi.org/10.1155/2015/986574

Reid, J. N., & Jamieson, R. K. (2022). A computational model of item-based directed forgetting. *Canadian Journal of Experimental Psychology*, *76*(2), 75–86. https://doi.org/10.1037/cep0000281

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303–345. https://doi.org/10.1111/j.1756-8765.2010.01111.x

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT Press. https://doi.org/10.7551/mitpress/6161.001.0001

Rumelhart, D. E. (1990). Brain style computation: Learning and generalization. In S. F. Zornetzer, J. L. Davis, & C. Lau (Eds.), *An introduction to neural and electronic networks* (pp. 405–420). Academic Press.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 3–30). MIT Press.

Sanders, C. A., & Nosofsky, R. M. (2020). Training deep networks to construct a psychological feature space for a natural-object category domain. *Computational Brain & Behavior*, *3*(3), 229–251. https://doi.org/10.1007/s42113-020-00073-z

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Senaldi, M. S. G., Titone, D. A., & Johns, B. T. (2022). Determining the importance of frequency and contextual diversity in the lexical organization of multiword expressions. *Canadian Journal of Experimental Psychology*, *76*(2), 87–98. https://doi.org/10.1037/cep0000271

Shaoul, C., & Westbury, C. (2010). *The westbury lab wikipedia corpus 2010*. University of Alberta.

Shiffrin, R. M. (2010). Perspectives on modeling in cognitive science. *Topics in Cognitive Science*, *2*(4), 736–750. https://doi.org/10.1111/j.1756-8765.2010.01092.x

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical bayesian methods. *Cognitive Science*, *32*(8), 1248–1284. https://doi.org/10.1080/03640210802414826

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Simon, H. A. (1969). *The sciences of the artificial*. MIT Press.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*(1), 49–65. https://doi.org/10.1016/S0896-6273(00)80821-1

Taler, V., & Johns, B. (2022). Using big data to understand bilingual performance in semantic fluency: Findings from the Canadian Longitudinal Study on Aging. *PLOS ONE*, *17*(11), Article e0277660. https://doi.org/10.1371/journal.pone.0277660

Taler, V., Johns, B. T., & Jones, M. N. (2020). A large scale semantic analysis of verbal fluency across the aging spectrum: Data from the Canadian longitudinal study on aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *75*(9), e221–e230. https://doi.org/10.1093/geronb/gbz003

Taler, V., Johns, B. T., Young, K., Sheppard, C., & Jones, M. N. (2013). A computational analysis of semantic structure in bilingual fluency. *Journal of Memory and Language*, *69*(4), 607–618. https://doi.org/10.1016/j.jml.2013.08.004

Taler, V., & Phillips, N. A. (2008). Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology*, *30*(5), 501–556. https://doi.org/10.1080/13803390701550128

Thiessen, E. D., & Pavlik, P. I., Jr. (2013). iMinerva: A mathematical model of distributional statistical learning. *Cognitive Science*, *37*(2), 310–343. https://doi.org/10.1111/cogs.12011

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, *4*(10), 1029–1038. https://doi.org/10.1038/s41562-020-0924-8

Tiv, M., Kutlu, E., Gullifer, J. W., Feng, R. Y., Doucerain, M. M., & Titone, D. A. (2022). Bridging interpersonal and ecological dynamics of cognition through a systems framework of bilingualism. *Journal of Experimental Psychology: General*, *151*(9), 2128–2143. https://doi.org/10.1037/xge0001174

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Wingfield, C., & Connell, L. (2022). Understanding the role of linguistic distributional knowledge in cognition. *Language, Cognition and Neuroscience*, *37*(10), 1220–1270. https://doi.org/10.1080/23273798.2022.2069278

Wulff, D. U., De Deyne, S., Jones, M. N., Mata, R., & the Aging Lexicon Consortium. (2019). New perspectives on the aging lexicon. *Trends in Cognitive Sciences*, *23*(8), 686–698. https://doi.org/10.1016/j.tics.2019.05.003

Zou, W., & Bhatia, S. (2021). Judgment errors in naturalistic numerical estimation. *Cognition*, *211*, Article 104647. https://doi.org/10.1016/j.cognition.2021.104647

Received August 10, 2022
Revision received November 23, 2022
Accepted December 7, 2022 ■