Instance theory as a domain-general framework for cognitive psychology

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

Abstract | The dominant view in cognitive psychology is that memory includes several distinct and separate systems including episodic memory, semantic memory and associative learning, each with a different set of representations, explanatory principles and mechanisms. In opposition to that trend, there is a renewed effort to reconcile those distinctions in favour of a cohesive and integrative account of memory. According to instance theory, humans store individual experiences in episodic memory and general-level and semantic knowledge such as categories, word meanings and associations emerge during retrieval. In this Perspective, we review applications of instance theory from the domains of remembering, language and associative learning. We conclude that instance theory is a productive candidate for a general theory of cognition and we propose avenues for future work that extends instance theory into the domain of cognitive computing, builds hybrid instance models and builds bridges to cognitive neuroscience.

The goal of psychology is to generate a coherent explanation of behaviour that generalizes over problems and domains[1]. Yet the discipline has adopted a divide and conquer strategy with distinct research groups focused on explaining different phenomena[2]. This situation has led to a divisive view of memory composed of distinct theories for different systems including episodic memory, semantic memory, procedural memory, priming, conditioning and non-associative learning[3]. Based on that perspective, it has been asserted that identifying general principles of memory is a fantasy and that no profound generalizations can be made about memory as a whole[4,5].

Modern psychology often distinguishes between explanations of memory of the specific and knowledge of the general. In the field of memory, the distinction is discussed as episodic versus semantic[6]. In categorization, the distinction is discussed as exemplars versus prototypes[7]. In language, the distinction is discussed as instances versus rules[8,9]. In learning, the distinction is discussed as memory versus association[10,11]. For some time, this divide and conquer strategy bore fruit, producing increasingly articulate theories albeit for increasingly specific laboratory behaviours. However, cognitive psychology has become a collection of disconnected explanations for behaviours in particular laboratory experiments. The strategic division of problems has been reified and the description and understanding of cognition had become something akin to a boutique Swiss army knife — a collection of independently useful tools grown too fat to fit into anyone's pocket.

However, a countermovement has proposed integration over division. For example, Newell famously argued that psychology had been seduced into playing a losing game with nature, not unlike the parlour game 'Twenty Questions'[2]. He suggested that researchers had grown accustomed to testing increasingly fine binary oppositions, operating under the illusion that the strategy would eventually whittle nature down to an indivisible truth. Newell pointed out that the strategy was flawed and recommended that the field take on the productive (albeit more difficult) problem of building an integrative and computationally articulate theory of cognition — the hallmark of a mature science. Unfortunately, Newell's warning has been largely ignored and the divisionist strategy remains in place[1,12].

Over the past 45 years, computational modelling has been at the forefront of developing an integrative theory of cognitive psychology. The computational view of cognition regards the taxonomy of independent memory systems as a theoretical challenge in need of a cohesive explanatory framework. There are now several integrative models of memory and cognition that together present a picture of theoretical progress that knits together data and behaviours over an impressive range of cognitive domains, including memory[13–19], attention[20–22], categorization[7,13,20,23], action[24,25], decision-making[26,27], language[28–32] and associative learning[33–37]. However, that programme remains in progress rather than completed.

In this Perspective, we argue that instance theory as represented in instance models (such as MINERVA[13,38]) is an articulate, coherent and productive framework to integrate data and theory across different branches of cognitive psychology. To make that case, we review a shortlist of successful applications of an integrated mechanistic framework to memory, language and associative learning. We conclude that instance theory is a capable and powerful organizing framework to guide cognitive research that can advance modern pursuits in big data and cognitive neuroscience.

Instance theory

Instance theory is a theoretical perspective that assumes memory is a single system that records specific events, or instances. When a retrieval cue (or probe) is presented (such as a word or picture), the cue interacts with stored information to create the memory of a previously experienced event. Unlike multiple system accounts of memory, general knowledge (such as category and conceptual knowledge) is not represented directly in a semantic memory system but, rather, emerges from episodic memory during retrieval[39–41].

There are several computational instance models that serve as formal instantiations of instance theory. The context model[7], resonance theory[42] and MINERVA[13,43,44] were amongst the first. However, the list has grown to include the generalized context model[20,45,46], the exemplar-based random walk model[23], instance theory of attention and memory[21,22], the knowledge model[24,25], the retrieving effectively from memory model[16] and the MINERVA family of models (that we refer to collectively as MINERVA in the remainder of this Perspective)[26,27,30,31,33,35,36,47]. Although these models differ in their computational details, they all share the perspective that memory records specific experiences and that general knowledge emerges 'on the fly' during retrieval from that store of specific experience.

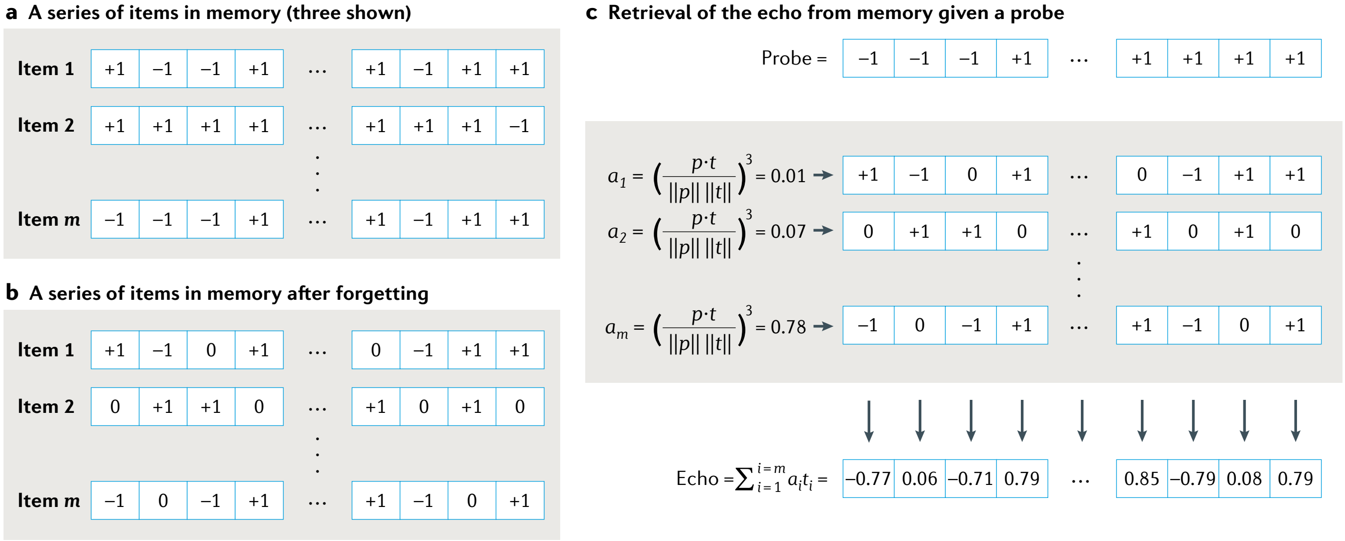
MINERVA is an instance model that articulates representation, storage and retrieval from memory (FIG. 1). According to the model, each experience is stored to memory as a unique trace. Presenting a retrieval cue causes each trace to become activated in proportion to its similarity to the cue (the word 'doctor' activates memory of having studied the word 'nurse' more than it activates memory of having studied the word 'dog'), and retrieval produces a weighted sum of the activated memory traces, known as the echo. Because traces record all events of a trial and whole traces are retrieved from memory, a retrieval cue also retrieves memory for events it has co-occurred with in the past; this is how the model simulates cued recall, associative learning and categorization. The information in the echo is assessed relative to the question at hand. In recognition, a retrieval cue that successfully retrieves itself (assessed by comparing the echo with the retrieval probe) is considered to be recognized. In semantic word meaning, the echo that is retrieved stands for the word's meaning. In associative learning, the echo that a cue retrieves is assessed for the presence of an associate. Although we focus our narrative on MINERVA owing to its particularly broad history of application[13,26,27,31–36,38,43,44,48–62], we will highlight where other instance models converge to reinforce our principal thesis that instance theory broadly is a coherent general framework for integrating and understanding cognition.

Memory and categorization

Instance theory was created to challenge the assertion that memory is divided into distinct episodic and semantic memory systems. Although there are many examples in that debate, we focus on the examples of prototype abstraction, the Deese–Roediger–McDermott (DRM) false memory effect, and dissociations of episodic and semantic memory in amnesia.

Prototype abstraction. In a prototype abstraction task, participants study category exemplars that are derived as random perturbations of an unpresented prototype (for example, dot patterns). At test, people are best at categorizing exemplars that they studied but are also better at categorizing the unstudied prototype than novel unstudied exemplars. Following a long delay between study and test, better categorization of the unstudied prototype (compared with novel unstudied exemplars) grows even stronger. Based on these results, it has been argued that people abstract and store the prototype in semantic memory, using it to support categorization judgements. It is further assumed that specific memories of studied items in episodic memory are forgotten faster than information in semantic memory[63,64].

Although people's judgements in the prototype abstraction task are consistent with semantic abstraction and a division of memory into separate episodic and semantic systems, the data pattern is also consistent with a single memory system account. According to instance theory, people encode category exemplars presented at study into a single memory store. At test, they judge each item based on its aggregate similarity to the stored category exemplars. Thus, people categorize studied exemplars as best, the unstudied prototype second-best (owing to its partial similarity to each item in the study list) and new items worst. Because forgetting over a study–test delay causes items in memory to grow increasingly average (less distinct from each other), instance theory predicts that following a delay, people's categorization of the items they studied should weaken and their categorization of the unpresented prototype



a A series of items in memory (three shown)
Item 1 +1 −1 −1 +1 … +1 −1 +1 +1
Item 2 +1 +1 +1 +1 … +1 +1 +1 −1
Item m −1 −1 −1 +1 … +1 −1 +1 +1
b A series of items in memory after forgetting
Item 1 +1 −1 0 +1 … 0 −1 +1 +1
Item 2 0 +1 +1 0 … +1 0 +1 0
Item m −1 0 −1 +1 … +1 −1 0 +1
c Retrieval of the echo from memory given a probe
Probe = −1 −1 −1 +1 … +1 +1 +1 +1
$a_1 = \left(\frac{p \cdot t}{\|p\|\,\|t\|}\right)^3 = 0.01$ → +1 −1 0 +1 … 0 −1 +1 +1
$a_2 = \left(\frac{p \cdot t}{\|p\|\,\|t\|}\right)^3 = 0.07$ → 0 +1 +1 0 … +1 0 +1 0
$a_m = \left(\frac{p \cdot t}{\|p\|\,\|t\|}\right)^3 = 0.78$ → −1 0 −1 +1 … +1 −1 0 +1
Echo = $\sum_{i=1}^{i=m} a_i t_i$ = −0.77 0.06 −0.71 0.79 … 0.85 −0.79 0.08 0.79

Fig. 1 | **The MINERVA model of memory. a** | Series of item traces (1 through m) stored to a memory matrix. Each item is represented as a row vector with each feature having a value of 1 or −1. **b** | Memory for items after forgetting (zeroes replace some of the elements). **c** | The retrieval cue (probe) activates each trace in proportion to its similarity to the cue. a_i values indicate activation of each trace in memory. Activation is computed as the normalized dot product of a probe and trace, raised to exponent 3. Activation values range between −1 and 1; 1 indicates probe and trace are identical, 0 indicates they are orthogonal and −1 indicates they are opposite. Thus, trace m with $a_m = 0.78$ is very similar to the probe, whereas trace 1 with $a_1 = 0.01$ is nearly orthogonal. Information retrieved from memory is called the echo and is an activation-weighted sum of traces in memory, where each trace contributes to that sum in proportion to its activation by the probe ($a_i \times t_i$). Thus, traces most similar to the retrieval cue (those with higher activation values) are represented more prominently in the echo.

a Representation of the category and five category exemplars

Category prototype =	1	1	1	1	1	−1	−1	−1	−1	−1
Category exemplars =	1	−1	1	1	−1	−1	−1	−1	−1	−1
	1	1	−1	1	1	−1	−1	−1	−1	−1
	1	1	1	−1	−1	−1	−1	−1	−1	−1
	1	1	1	1	1	1	−1	−1	−1	1
	1	1	1	1	1	−1	−1	−1	1	−1

b The five category exemplars as traces, *t*, stored to memory with some forgetting

t_1 =	0	−1	1	1	−1	−1	−1	−1	−1	−1
t_2 =	1	1	−1	1	0	−1	−1	−1	−1	0
t_3 =	1	1	1	−1	−1	0	−1	0	−1	−1
t_4 =	1	1	1	1	0	1	0	−1	−1	0
t_5 =	1	1	1	1	1	−1	−1	−1	0	−1

c A novel category test probe, *p*, based on the category prototype

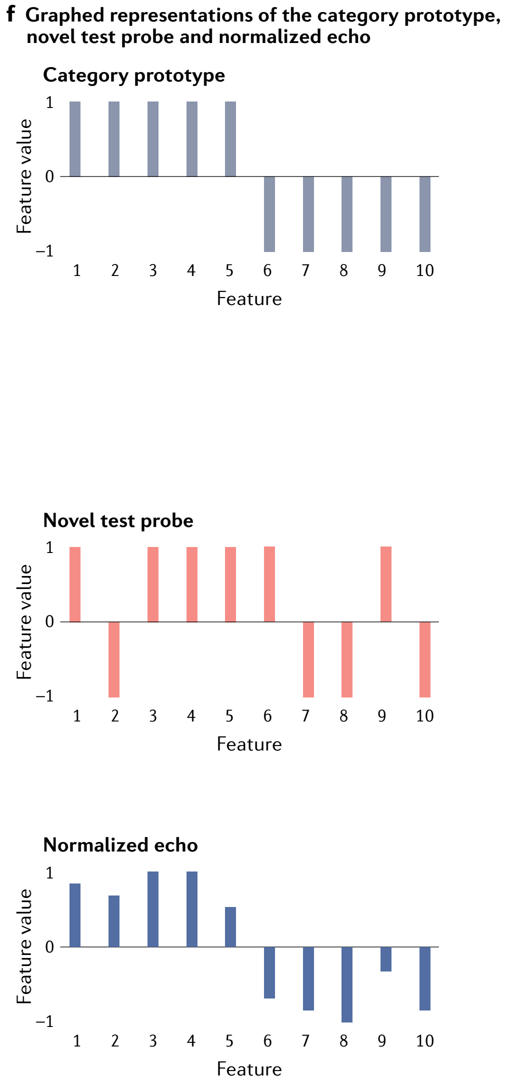
p =	1	−1	1	1	1	1	−1	−1	1	−1

d Memory, *M*, where each trace is activated in proportion to its similarity to the probe, *p*

a_1 = 0.03 →	0.00	−0.03	0.03	0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03
a_2 = 0.00 →	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
a_3 = 0.00 →	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
a_4 = 0.03 →	0.03	0.03	0.03	0.03	0.00	0.03	0.00	−0.03	−0.03	0.00
a_5 = 0.13 →	0.13	0.13	0.13	0.13	0.13	−0.13	−0.13	−0.13	0.00	−0.13

e The echo, *e*, retrieved by the novel test probe, *p* (normalized echo = *e*′)

$e = \sum_{i=1}^{i=m} a_i t_i$ =	0.16	0.13	0.19	0.19	0.10	−0.13	−0.16	−0.19	−0.06	−0.16
$e' = e/\max(\lvert e \rvert)$ =	0.84	0.68	1.00	1.00	0.53	−0.68	−0.84	−1.00	−0.32	−0.84



f Graphed representations of the category prototype, novel test probe and normalized echo

Fig. 2 | **'On the fly' memory retrieval in MINERVA.** According to instance theory, humans store specific experiences (instances) but not general knowledge in memory. Consider how retrieval is implemented in MINERVA[13]. **a** | Representations for a category prototype and five category exemplars derived from the prototype. **b** | These five exemplars stored to a memory matrix with some information loss (some values turned to 0). **c** | Novel test probe not in the study list but also a category exemplar. **d** | Memory matrix where each trace is activated by its similarity to the novel probe (item 5 is activated most strongly). **e** | Echo retrieved by the test probe and normalized version of the echo. **f** | Category prototype, novel test probe and normalized echo presented as bar graphs. The novel test exemplar retrieves an echo that resembles the category prototype. In fact, the echo is more similar to the category prototype than to the novel test probe. This example demonstrates how MINERVA retrieves a category prototype 'on the fly' from memory of studied category exemplars, even when the prototype is not directly represented in memory.

should remain stable or even strengthen (FIG. 2). Simulations with MINERVA confirm this explanation[13,44,57], as do simulations with other instance models including the context model, the generalized context model and the exemplar-based random walk model[7,23,65].

DRM false memory effect. The DRM false memory effect represents a linguistic twist on prototype abstraction[66,67]. This effect is observed when people falsely recognize a critical unstudied word such as 'sleep' that is associated with a thematically related study list such as 'rest', 'bed', 'nap' and 'blanket'. The dual-process explanation for the DRM effect is that people encode the studied words in episodic memory and encode an abstracted representation of the list's gist (its average meaning) in semantic memory[68]. At test, people correctly recognize the studied items based on their similarity to the studied items in episodic memory and false alarm to the critical word based on its match to the gist stored in semantic memory — a phenomenon interpreted more generally as a false memory effect.

By contrast, according to instance theory, people encode each studied word in episodic memory. At test, people recognize studied items based on a specific match to representations in memory. On this account, people misrecognize a critical unstudied word ('sleep') not owing to a gist stored in semantic memory but because of its combined partial match to all items in memory ('rest', 'bed', 'nap' and 'blanket'). This line of reasoning has been confirmed in simulations with MINERVA[50,69,70], confirming that instance theory predicts the DRM false memory effect without requiring a separate semantic memory system. There are several related false memory effects, such as false memory for inferences that people make while reading[71,72] and schema effects[73], that can also be explained in the same manner, within a single system account of memory.

Memory dissociations. Data from memory impairment provide another source of evidence for the purported semantic–episodic memory distinction. Individuals with amnesia display more severe impairments on episodic memory tasks (recognition) than semantic memory tasks (categorization)[74,75]. Based on that dissociation, it has been argued that amnesia reflects a selective impairment to a distinct episodic memory system without corresponding impairment to a distinct semantic memory system. However, a growing body of evidence has begun to question this view[76–78]. As an example, consider a now classic experiment testing the dissociation between categorization and recognition[74]. In the experiment, individuals with amnesia and control participants studied dot patterns derived from a prototype (as in the prototype abstraction task). Following study, they performed categorization and recognition tasks for studied and unstudied patterns. Whereas the individuals with amnesia performed as well as control participants at categorization, they performed worse than control participants at recognition. To explain the dissociation, a dual systems theory assumes that amnesia represents impaired episodic memory coupled with spared semantic memory.

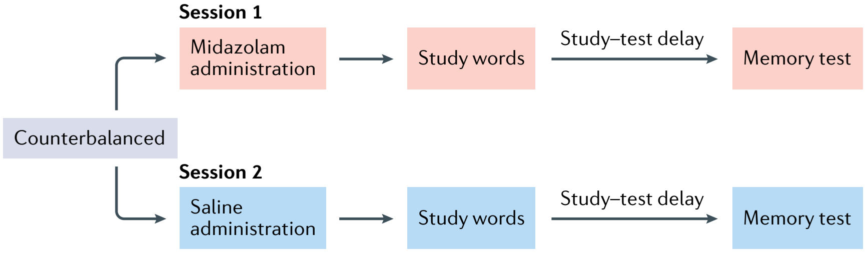
This result seems to contradict an instance view: if recognition and categorization are both based upon memory of studied exemplars, then forgetting the exemplars should impair both categorization and recognition equally. However, the pattern can be successfully simulated with instance theory. As with prototype abstraction and the DRM effect, poor memory of the exemplars renders them increasingly average (less distinct). This change makes the items in memory more difficult to distinguish from one another, reducing successful recognition of specific studied items without reducing successful categorization of items that are similar to the studied set. Thus, if amnesia results in poor memory for studied exemplars, instance theory predicts a far more severe impairment in recognition of the studied exemplars than categorization. This prediction and related predictions have been confirmed using several instance models including MINERVA[49,57,79,80], the general context model[65,81–83], the neural exemplar theory[84] and the retrieving effectively from memory model[85] (BOX 1).

Box 1 | **Modelling the nature of encoding difficulties in amnesia**

In several experiments, humans' memory for presented words was tested following administration of midazolam — a drug used to induce transient amnesia during surgery — versus a control condition. Over those studies, the order of drug versus control administration was randomized so that some participants were tested under the drug condition before the control, and vice versa (see the figure). Each participant completed both the drug and saline conditions to ensure a high precision and within-subject measurement of the drug effect[157–161].

These experiments revealed several dissociations in recognition memory performance between the induced amnestic state and the control condition. Notably, a reversal of the standard recognition advantage for low-frequency relative to high-frequency words was found, which was interpreted as evidence for a dual-process account of recognition memory[162]. However, work using the retrieving effectively from memory model demonstrated that this reversal is also consistent with a single-process instance theory[16,85]. According to the analysis and consistent with explanations of memory dissociations with the MINERVA and generalized context models[49,57,81], midazolam decreases the accuracy with which memory traces are stored.

The retrieving effectively from memory model is unique among instance models in that it includes two encoding parameters. One parameter controls the probability of encoding features to memory. The other controls the probability of encoding a feature correctly, given that it is encoded. Based on a simulation analysis, the authors concluded that performance under midazolam is consistent with noisier rather than less encoding of study items[85]. The distinction between encoding and encoding correctly is important for understanding how amnesia impacts memory, enabling a more precise understanding of how human memory operates. This understanding, in turn, can inform the design of behavioural treatments and strategies that people with amnesia can use to compensate for and ameliorate their memory difficulties.



Taken together, our discussion of prototype abstraction, the DRM effect and amnesia demonstrates that instance theory can re-explain so-called semantic memory effects as emergent artefacts of retrieval from episodic memory. Thus, phenomena traditionally thought to require a theoretical division of memory into separate episodic and semantic systems can, instead, be understood as reflecting a common set of cognitive representations, principles, and operations in a single memory system.

Language

The study of language has long been dominated by the nativist or universalist perspective that language is a specialized system operating independently from general cognition[86]. However, that perspective has been challenged by usage-based accounts that language emerges as an interaction of general cognitive functions and the language environment[8]. Instance theory has served as a formal framework to test the usage-based hypothesis, in the domains of word meaning and speech normalization[30,54,87,88].

Word meaning. Virtually all accounts of word meaning in semantic memory assume that a word's meaning is represented by a lexical entry averaging over its history of use[89]. Consistent with this prototype perspective, computational accounts of semantic memory such as word2vec (REF.[90]) or BEAGLE[29] encode a word's meaning as a single vector that averages over the word's usage in printed text[28,29,90–92]. Once computed, these stored lexical vectors (prototypes) in semantic memory accurately predict people's judgements about word meanings[93–96]. However, these explanations require a special abstraction process across each word's history and a division of memory into separate semantic and episodic systems.

Two instance-based models of 'semantic memory' have been developed within the MINERVA framework: the instance theory of semantics[30] and its predecessor, the constructed semantics model[31]. In these models, individual language experiences are encoded as traces in episodic memory and word meanings are constructed during retrieval as the echo computed from the episodic store. In the instance theory of semantics, each word is represented by a unique vector and each language experience (for example, a paragraph in a newspaper) is represented as a sum of vectors that correspond to the words in the experience. When a word is presented as a retrieval cue

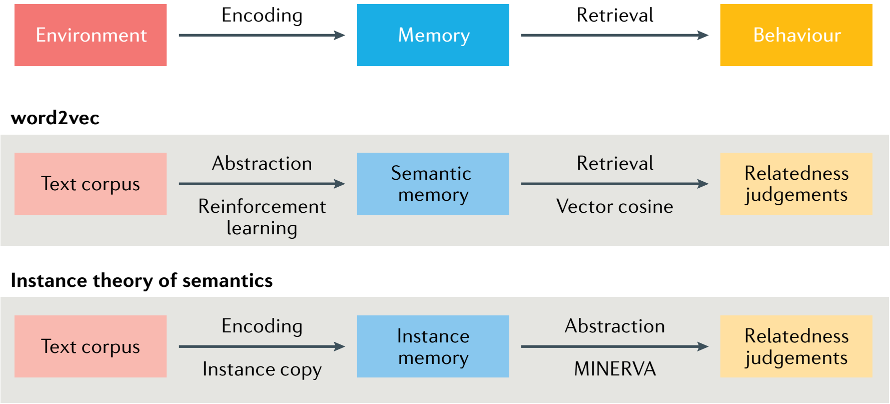


Fig. 3 | **Storage versus retrieval accounts of semantic memory in language.** The generally accepted schematic of memory consists of encoding, storage and retrieval. By contrast, instance theory of semantics[30] and the popular word2vec model[90] learn from a text corpus and can accurately simulate semantic relatedness judgements. word2vec does so by abstracting a representation for later retrieval; it posits semantic abstraction at encoding and storage of a single semantic representation per word. Instance theory of semantics simulates relatedness judgements by retrieving representations as needed. Instance theory of semantics has no semantic memory — the behaviour that makes it appear that it has semantic memory emerges from the retrieval process.

to memory, it activates each memory trace, and the weighted sum of the traces — the echo — stands for the word's meaning.

Because word meanings are constructed during retrieval each time a word is encountered, instance models naturally disambiguate the intended meaning of homonyms in context because the activated traces in memory are those most similar to the current context. By contrast, prototype models collapse all distinct meanings of a word into a single aggregate representation — a scenario that challenges disambiguation of competing meanings. For example, prototype models often have difficulty switching between the frequent meaning of 'bank' cued by the sentence 'I withdrew money from the bank' to one of its less frequent meanings as indicated in the sentence 'I canoed to the bank'[97], although there has been recent progress on the disambiguation of word meaning in these models[98]. Instance theory finesses the problem by constructing the meaning for 'bank' from differential activation of exemplars that best match how it is being used in the moment[30,59,89].

Despite the differences in how prototype models and instance models arrive at word meanings, whether prospectively by prototype abstraction or retrospectively by memory retrieval, they arrive at similar conclusions regarding word meaning. However, despite arriving at similar conclusions (FIG. 3), the instance theory of semantics provides an existence proof that a separate semantic abstraction process coupled with storage in a separate semantic memory system is unnecessary to understand how people know word meanings. This model provides the kind of computational framework that is needed to bring the reconsideration of separate episodic and semantic memory systems into clear analytic focus.

Speech normalization. Speech is a noisy signal; each utterance bears idiosyncrasies due to speaker variability, prosody and regional accent[99]. For example, 'about' and 'sorry' are pronounced differently when spoken by someone with a Canadian versus an American accent. Yet people are remarkably adept at recognizing and comprehending spoken words despite these variations.

According to the standard view, speech variability is solved by an early perceptual process that recognizes spoken words and normalizes them for use in retrieval from semantic memory[100,101]. However, such a normalization process conflicts with the fact that people's memory performance is impacted by differences in word pronunciation between study and test lists. For example, people recognize studied words less well when different people speak those words at study versus test, suggesting that memory for spoken words retains their auditory details[51,102,103].

By contrast to semantic memory theories, instance theory proposes that speech is remembered without undergoing normalization and that speech normalization is a side effect of retrieval from episodic memory. According to this account, memory traces of prior speech experiences retain their auditory idiosyncrasies such as speaker identity and accent. When a new speech utterance is heard, it activates memory of all similar speech instances and a weighted average of the activated memory traces is retrieved. Because the information retrieved is a weighted average, that information emphasizes features that are shared by the activated traces and washes out their differences. By that process, presenting a word retrieves its normalized form.

Simulations with MINERVA demonstrated that an instance model accounts for people's behaviour in speech shadowing and speech recognition experiments[32]. Follow-up work demonstrated that the model also accounts for complicated patterns in lexical decision (deciding whether an utterance is a real or made-up word)[59]. In contrast to the view that speech normalization is an early perceptual process, the instance view explains normalization as a corollary of an instance-based retrieval process.

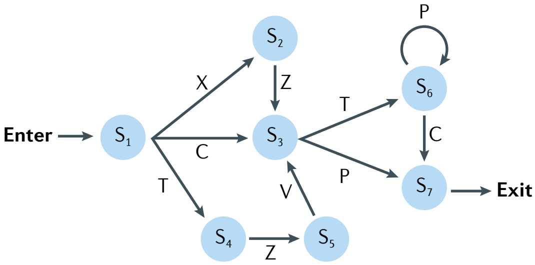
Instance theory provides a coherent explanation of (at least) some aspects of language cognition. We have focused on word meaning and speech normalization but instance theory has also been applied to other language problems including sentence production[51], statistical learning[104] and implicit learning of grammatical structure (BOX 2). Based on those demonstrations, and consistent with the usage-based account of language[8], this work strongly suggests that language perception and comprehension can be understood as an interaction between language experience and an instance-based approach to memory.

Learning

Associative learning refers to the study of the basic cognitive processes that govern behaviour of human and non-human animals. Yet models of associative learning and models of memory have been developed independently. Instance theory has had an important role in a productive reintegration of data and theory in the two domains[10,33,34,37,105,106].

Associative learning. In a simple associative learning procedure, a cue (such as an auditory tone) is presented followed by an outcome (such as food). With experience, presentation of the cue elicits behavioural anticipation of the outcome (such as salivation or a lever press); subsequent presentation of the cue in the absence of the outcome produces extinction (weakening) of the learned anticipation. The wax and wane of cue-driven anticipation is associative learning.

Box 2 | **Implicit learning of grammatical structure**



Researchers have used the artificial-grammar task to examine the phenomenon that humans learn the regularities in their environment without explicit intent[40,163,164]. In this task, participants memorize letter strings constructed according to the rules of an artificial grammar that specifies which letters can follow one another from left to right. A grammatical stimulus is generated by starting at the leftmost node of a grammar (marked S_1 in the figure) and following the paths (indicated by arrows) until reaching the exit node marked S_7. Following the grammar in the figure, participants might be asked to memorize the strings CTPC, TZVTPC, XZTPPPPC and CTC. In a test phase, they are told the strings they just memorized were constructed according to rules of an artificial grammar and they are invited to discriminate unstudied grammatical strings (such as CTPPPC and TZVP) from unstudied and ungrammatical strings (such as VTTTPX and TZVZTPC). Although participants can discriminate the two kinds of items above chance, they cannot describe the underlying grammar.

The standard account of performance in artificial-grammar learning tasks is that a specialized implicit learning system abstracts the grammar during study and applies that knowledge at test. Because participants cannot describe the grammar, it is assumed that the knowledge and use of the grammar is unconscious. By contrast, instance theory proposes that performance can be explained without invoking a specialized system. MINERVA was applied to the artificial-grammar task by encoding study strings to memory and then judging test strings based on how well they can be reconstructed during retrieval[52]. The model distinguished grammatical from ungrammatical test strings without knowing the grammar — just as humans can. The model's ability to perform discrimination follows from the natural correlation between the form and amount of structure in an instance produced from a grammar[165], and the structure of the grammar itself combined with the fact that the information retrieved from memory aggregates over the individual items.

Thus, performance in the artificial-grammar task can be understood using the same principles and mechanisms that have long been used to understand retrieval from episodic memory. Although we have highlighted the use of MINERVA in this domain, other researchers have used the generalized context model to make the same point[166]. The convergence on a common conclusion using different instance models reinforces our principal thesis that instance theory, in general, can explain examples of implicit language learning[166].

Associative learning has traditionally been modelled as a current summary of associative strengths between cues and outcomes. These representations are represented as current summaries over an entire learning history and independent of the learner's memory for the individual trials that contributed to the current summary[107]. However, the position that an association is represented separately and independently from memories of individual learning trials is difficult to reconcile with the body of evidence that animals, human and non-human, have an impressive ability to remember the events of individual learning trials[108–116].

By contrast to classical learning theories, instance theory identifies the individual experience — the instance — as the primitive unit of knowledge and models associative learning as the storage and retrieval of instances from memory. From that perspective, association is a corollary of retrieval from memory rather than a summary over the learning history. To test instance theory in this context, MINERVA was applied to associative learning protocols[33,34]. The theory was augmented to account for the fact that animals encode unanticipated and, therefore, surprising events more strongly than anticipated events (surprise-driven encoding)[117,118]. With surprise-driven encoding, MINERVA predicted a broad range of associative learning phenomena.

The most compelling demonstration of the power of instance models is an explanation of retrospective revaluation — a learning phenomenon that evades easy explanation by classical theories of associative learning[119–122]. In retrospective revaluation, participants might learn, for example, that apples and bananas together cause an allergic reaction. Accordingly, the learner might conclude that apples and bananas are equally responsible for the reaction. However, if they later learn that apples alone cause the allergic reaction, they might update their assertion that only apples cause the allergy and, therefore, bananas do not. This phenomenon requires memory of prior trials to retrospectively revaluate associations and assertions in light of new information. MINERVA enables retrospection as a natural outcome of being a memory theory because remembering prior events supports the capacity to rethink their relationships to the outcome. By contrast, classical learning theories require ad hoc assumptions to enable appropriate retrospective revaluation in the absence of memory for prior trials[122,123].

Evaluative conditioning. Evaluative conditioning is an example of associative learning that denotes how a person's attitude changes towards a conditioned stimulus based on its pairing with an emotionally valenced unconditioned stimulus[124,125]. For example, seeing a company logo paired with a smiling actor may lead to more positive evaluations of the logo. Evaluative conditioning has attracted a good deal of applied interest in relation to advertising and advertising ethics[126]. However, it has also attracted interest as a theoretical conundrum.

According to classical associative learning theory, a person's learned reaction to a conditioned stimulus is really a reaction to its anticipated unconditioned stimulus. Thus, if the conditioned stimulus–unconditioned stimulus association is extinguished, it is expected that the conditioned stimulus should cease to elicit the reaction. However, evaluative conditioning represents a situation in which the reaction to the conditioned stimulus persists even after the conditioned stimulus–unconditioned stimulus relationship is extinguished. Based on the persistence of evaluative conditioning despite extinction of the conditioned stimulus–unconditioned stimulus association, it has been argued that evaluative conditioning represents a special form of associative learning.

Evaluative conditioning can also be explained within a single system instance-based account. According to the instance theory explanation, people's reaction to a conditioned stimulus reflects retrieval of memories from both the acquisition and extinction phases of an evaluative conditioning experiment whereas their reaction to an unconditioned stimulus reflects retrieval from the more recent extinction phase. In a simulation with MINERVA, memory for a trial included a contextual time stamp that distinguished when the trial was encountered; during either the acquisition or the extinction phase of the training protocol[35,36]. At test, the context was included as part of the probe so the model retrieved selectively from one or both of the training phases[127]. Given this context-selective retrieval, MINERVA captured the persistence of evaluative conditioning following extinction of the conditioned stimulus–unconditioned stimulus association[106]. Thus, if one acknowledges context-sensitive retrieval from memory differences in the evaluations of the conditioned stimulus relative to the conditioned stimulus–unconditioned stimulus association, expectancy can be understood as reflecting different context-sensitive summaries of the same learning history[128,129]. Accordingly, the difference does not require different learning processes for evaluative conditioning and associative learning[10,105].

These examples demonstrate that instance theory captures multiple examples of associative learning and presents a promising avenue for investigating learning behaviour in both human and non-human animals. Instance theory therefore supports an integrated understanding of the mechanisms that support learning and memory.

Conclusion

Instance theory is a coherent approach to understanding cognition. It shifts the complexity of memory from encoding to retrieval such that cognitive complexity does not reflect a battery of sophisticated encoding mechanisms but, instead, emerges as a corollary of constructive retrieval from episodic memory. In focusing on memory, language and learning, we have not discussed applications of instance theory in other domains such as attention[21,22], action and procedural memory[24,25], decision-making[26,27,41], choice reaction time[23,53] and implicit learning[40,52,56]. To the extent that a single system perspective is successful, the need for multiple parallel theories to explain these different phenomena can be scrutinized as theoretical overreach.

The success of instance theory as an integrative explanation across domains shows that single system accounts are tenable. Thus, researchers should consider the necessity and veracity of a multiple system account of memory. Whether instance theory turns out to be the best integrative theory of cognition, the continued effort to test the necessity of divisions moves cognitive psychology towards true explanations for cognition. Instance theory can provide a governing framework to unify an understanding of human behaviour and motivate new experimental work and discoveries.

Instance models are often criticized for explaining cognition at the computational and algorithmic levels but not at the implementational (neuroscientific) level[130]. That distinction points out two clear avenues for research on instance theory. The first is to rewrite instance models in implementational terms. Such a project would serve to express instance theory in a common language to facilitate a more fluid exchange of data and ideas between psychology and cognitive neuroscience. A second, related avenue for future research is to rewrite instance models to be more computationally efficient. Whereas instance theory assumes that memory stores all individual experiences and that retrieval integrates over that entire history, implemented instance models must deal with the practical problem of exploding simulation times that grow with the number of instances in memory. Thus, if researchers want to apply instance models to big data problems such as language, it will be necessary to acknowledge and face the pragmatic complication of balancing computational efficiency and implementational form. Re-expressing instance models in terms of distributed representations would both deal with the computational efficiency of computing predictions and also resolve the tension between theoretically infinite memory and the capacity limits of the substrate in which memory resides. Promising frameworks for this programme include implementation-level instantiations of instance theory such as holographic representation[131–135], artificial neural networks[48,136–138] and the neural exemplar theory[84].

Instance theories take an extreme position by explaining all aspects of cognition without recourse to other forms of processing. As a result, they fall short of straightforward explanations for some classes of behaviour such as rule-based categorization. To bridge the gap, some researchers have constructed hybrid models that include an instance-based system operating in parallel with other forms of cognition, such as a rule-based classification system that discriminates category members according to one or more verbal if–then strategies[139–142]. Although these hybrid systems solve problems that instance theories have trouble negotiating, an interesting programme of research is to integrate the idea of rules and procedures of mind into an instance framework. For example, rules or processing operations might be stored into memories and enacted once retrieved. Doing so opens up questions and possibilities about how memories for actions and procedures might be represented within instance models to support storage and retrieval of behaviours, not just information. That idea would be consistent with skill and action-based work in instance theory[24,25,143] as well as Kolers' procedures of mind framework for cognition where memories encode procedures for behaviour rather than a record of stimulation[144]. Rosenbaum and colleagues' instance-based knowledge model provides a strong map to pursuing how instance models might be rewritten to predict how instance-based retrieval drives action.

Given the availability of large databases of text and image data, scaling up instance models to complicated and real-world domains is another promising and productive avenue for investigation. Pursuing that goal would present instance theory as a competitor framework in the domain of cognitive computing and applied cognitive science to modern and emerging technological perspectives on the design of cognitive systems[145]. As noted elsewhere[146,147], simple processes applied to complex data can yield surprisingly sophisticated cognitive behaviour. It should be an aim of ongoing research to investigate the scope and depth of sophistication that might emerge in the practice of applying instance models to big data sources.

We have focused, in large part, on explaining how instance theory provides an explanation of cognition over established examples of behaviour in cognitive psychology. However, whereas instance theory's success at postdicting what is already known is an important step, theories should also present opportunities for discovery. Instance theory might be profitably applied in this regard to the domain of ecological cognition[148,149]. This domain explores how people arrive at intuitive inferences and decisions that are irrational relative to prescriptive accounts of decision, yet rational relative to the ecology of the behavioural environment. This distinction has been explained by assuming that decision-making is tuned to a correspondence rather than coherence criterion[150]. Instance theory is an ideal framework for understanding the distinction between normative and descriptive accounts of decision-making, where it has already been applied to heuristics and normative reasoning[26,27,151]. Further applications of instance theory can be used to understand why people arrive at the illogical but well-structured and adaptive decisions that they do. For example, the instance theory of associative learning predicts people's decisions in reasoning experiments[26,27]. We are excited to examine this problem moving forward by integrating basic mechanisms of associative learning and memory with decision-making and intuitive inference.

We have focused on examples of instance theory using the MINERVA model. However, we could have built the same case using other instance models such as the generalized context model or the retrieving effectively from memory model. Also, whereas instance theories agree in principle, instance models differ in some meaningful ways. For example, the generalized context model represents stimuli using multidimensional scaling[152,153], the feature model incorporates assumptions about perceptual representation[15], the retrieving effectively from memory model uses Bayes' theorem to integrate the influence of expectation on memory[16,151] and the rule-plus-exception model provides a way to integrate exemplar theory with rule-based processes[139,140,154]. Thus, although testing and comparing different instance models is a productive strategy, it would also be profitable to consolidate different instance models. One route might involve deriving a single nested model to identify and sort through the assumptions and mechanisms that the models do and do not share[135,155]. Another route is to create a hybrid model that retains the differences between the various models and conduct an analysis of that hybrid model to identify and distinguish necessary from merely useful assumptions and mechanisms of the various instance models[139,141]. Despite uncertainties in how instance theory might be implemented in a shared programme, there is value in committing to a shared model that coordinates a mature, integrative and cumulative research programme[1,156].

Randall K. Jamieson[1]✉, *Brendan T. Johns*[2], *John R. Vokey*[3] *and Michael N. Jones*[4]

[1]Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada.

[2]Department of Psychology, McGill University, Montréal, Quebec, Canada.

[3]Department of Psychology, University of Lethbridge, Lethbridge, Alberta, Canada.

[4]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA.

✉*e-mail: randy.jamieson@umanitoba.ca*

https://doi.org/10.1038/s44159-022-00025-3

Published online: 09 February 2022

1. Oberauer, K. & Lewandowsky, S. Addressing the theory crisis in psychology. *Psychon. Bull. Rev.* **26**, 1596–1618 (2019).
2. Newell, A. in *Visual Information Processing* (ed. Chase, W. G.) 283–308 (Elsevier, 1973).
3. Squire, L. R. Memory systems of the brain: a brief history and current perspective. *Neurobiol. Learn. Mem.* **82**, 171–177 (2004).
4. Roediger, H. L. III Relativity of remembering: why the laws of memory vanished. *Annu. Rev. Psychol.* **59**, 225–254 (2008).
5. Tulving, E. Episodic memory: from mind to brain. *Annu. Rev. Psychol.* **53**, 1–25 (2002).
6. Tulving, E. How many memory systems are there? *Am. Psychol.* **40**, 385 (1985).
7. Medin, D. L. & Schaffer, M. M. Context theory of classification learning. *Psychol. Rev.* **85**, 207–238 (1978).
8. Abbot-Smith, K. & Tomasello, M. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Rev.* **23**, 275–290 (2006).
9. McAndrews, M. P. & Moscovitch, M. Rule-based and exemplar-based classification in artificial grammar learning. *Mem. Cognit.* **13**, 469–475 (1985).
10. Bouton, M. E. & Moody, E. W. Memory processes in classical conditioning. *Neurosci. Biobehav. Rev.* **28**, 663–674 (2004).
11. Shanks, D. R. Learning: from association to cognition. *Annu. Rev. Psychol.* **61**, 273–301 (2010).
12. Szollosi, A. & Donkin, C. Arrested theory development: the misguided distinction between exploratory and confirmatory research. *Perspect. Psychol. Sci.* **16**, 717–724 (2021).
13. Hintzman, D. L. 'Schema abstraction' in a multiple-trace memory model. *Psychol. Rev.* **93**, 411–428 (1986).
14. Brown, G. D. A., Neath, I. & Chater, N. A temporal ratio model of memory. *Psychol. Rev.* **114**, 539–576 (2007).
15. Nairne, J. S. A feature model of immediate memory. *Mem. Cognit.* **18**, 251–269 (1990).
16. Shiffrin, R. M. & Steyvers, M. A model for recognition memory: REM — retrieving effectively from memory. *Psychon. Bull. Rev.* **4**, 145–166 (1997).
17. Murdock, B. B. A theory for the storage and retrieval of item and associative information. *Psychol. Rev.* **89**, 609–626 (1982).
18. Murdock, B. B. A distributed memory model for serial-order information. *Psychol. Rev.* **90**, 316–338 (1983).
19. Murdock, B. B. Context and mediators in a theory of distributed associative memory (TODAM2). *Psychol. Rev.* **104**, 839–862 (1997).
20. Nosofsky, R. M. Attention, similarity, and the identification–categorization relationship. *J. Exp. Psychol. Gen.* **115**, 39–57 (1986).
21. Logan, G. D. An instance theory of attention and memory. *Psychol. Rev.* **109**, 376–400 (2002).
22. Logan, G. Toward an instance theory of automatization. *Psychol. Rev.* **95**, 492–527 (1987).
23. Nosofsky, R. M. & Palmeri, T. J. An exemplar-based random walk model of speeded classification. *Psychon. Bull. Rev.* **5**, 345–369 (1998).
24. Rosenbaum, D. A., Loukopoulos, L. D., Vaughan, J., Meulenbroek, R. G. J. & Engelbrecht, S. E. Planning reaches by evaluating stored postures. *Psychol. Rev.* **102**, 28–67 (1995).
25. Rosenbaum, D. A., Meulenbroek, R. J., Vaughan, J. & Jansen, C. Posture-based motion planning: applications to grasping. *Psychol. Rev.* **108**, 709–734 (2001).
26. Dougherty, M. R. P., Gettys, C. F. & Ogden, E. E. MINERVA-DM: a memory processes model for judgments of likelihood. *Psychol. Rev.* **106**, 180–209 (1999).
27. Thomas, R. P., Dougherty, M. R., Sprenger, A. M. & Harbison, J. I. Diagnostic hypothesis generation and human judgment. *Psychol. Rev.* **115**, 155–185 (2008).
28. Landauer, T. K. & Dumais, S. T. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* **104**, 211–240 (1997).
29. Jones, M. N. & Mewhort, D. J. K. Representing word meaning and order information in a composite holographic lexicon. *Psychol. Rev.* **114**, 1–37 (2007).
30. Jamieson, R. K., Avery, J. E., Johns, B. T. & Jones, M. N. An instance theory of semantic memory. *Comput. Brain Behav.* **1**, 119–136 (2018).
31. Kwantes, P. J. Using context to build semantics. *Psychon. Bull. Rev.* **12**, 703–710 (2005).
32. Goldinger, S. D. Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* **105**, 251–279 (1998).
33. Jamieson, R. K., Crump, M. J. C. & Hannah, S. D. An instance theory of associative learning. *Learn. Behav.* **40**, 61–82 (2012).
34. Jamieson, R. K., Hannah, S. D. & Crump, M. J. C. A memory-based account of retrospective revaluation. *Can. J. Exp. Psychol.* **64**, 153–164 (2010).
35. Aust, F., Haaf, J. M. & Stahl, C. A memory-based judgment account of expectancy-liking dissociations in evaluative conditioning. *J. Exp. Psychol. Learn. Mem. Cognit.* **45**, 417–439 (2019).
36. Stahl, C. & Aust, F. Evaluative conditioning as memory-based judgment. *SPB* **13**, e28589 (2018).
37. Blough, D. S. Context reinforcement degrades discriminative control: a memory approach. *J. Exp. Psychol. Anim. Behav. Process.* **24**, 185–199 (2010).
38. Hintzman, D. L. Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychol. Rev.* **95**, 528–551 (1988).
39. Brooks, L. R. in *Cognition and Categorization* (eds Rosch, E. & Lloyd, B.) 169–211 (Wiley, 1978).
40. Vokey, J. R. & Brooks, L. R. Salience of item knowledge in learning artificial grammars. *J. Exp. Psychol. Learn. Mem. Cogn.* **18**, 328 (1992).
41. Kahneman, D. & Miller, D. T. Norm theory: comparing reality to its alternatives. *Psychol. Rev.* **93**, 136–153 (1986).
42. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **85**, 59–108 (1978).
43. Hintzman, D. L. MINERVA 2: a simulation model of human memory. *Behav. Res. Methods Inst. Comput.* **16**, 96–101 (1984).
44. Hintzman, D. L. & Ludlam, G. Differential forgetting of prototypes and old instances: simulation by an exemplar-based classification model. *Mem. Cognit.* **8**, 378–382 (1980).
45. Nosofsky, R. M. Exemplar-based accounts of relations between classification, recognition, and typicality. *J. Exp. Psychol. Learn. Mem. Cogn.* **14**, 700–708 (1988).
46. Nosofsky, R. M. in *Formal Approaches in Categorization* (eds Pothos, E. M. & Wills, A. J.) 18–39 (Cambridge Univ. Press, 2011).
47. Collins, R. N., Milliken, B. & Jamieson, R. K. MINERVA-DE: an instance model of the deficient processing theory. *J. Mem. Lang.* **115**, 104151 (2020).
48. Hintzman, D. L. Human learning and memory: connections and dissociations. *Annu. Rev. Psychol.* **41**, 109–139 (1990).
49. Jamieson, R. K., Holmes, S. & Mewhort, D. J. K. Global similarity predicts dissociation of classification and recognition: evidence questioning the implicit–explicit learning distinction in amnesia. *J. Exp. Psychol. Learn. Mem. Cogn.* **36**, 1529–1535 (2010).
50. Arndt, J. & Hirshman, E. True and false recognition in MINERVA2: explanations from a global matching perspective. *J. Mem. Lang.* **39**, 371–391 (1998).
51. Goldinger, S. D. & Azuma, T. Episodic memory reflected in printed word naming. *Psychon. Bull. Rev.* **11**, 716–722 (2004).
52. Jamieson, R. K. & Mewhort, D. J. K. Applying an exemplar model to the artificial-grammar task: inferring grammaticality from similarity. *Q. J. Exp. Psychol.* **62**, 550–575 (2009).
53. Jamieson, R. K. & Mewhort, D. J. K. Applying an exemplar model to the serial reaction-time task: anticipating from experience. *Q. J. Exp. Psychol.* **62**, 1757–1783 (2009).
54. Johns, B. T., Jamieson, R. K., Crump, M. J. C., Jones, M. N. & Mewhort, D. J. K. Production without rules: using an instance memory model to exploit structure in natural language. *J. Mem. Lang.* **115**, 104165 (2020).
55. Jamieson, R. K., Mewhort, D. J. K. & Hockley, W. E. A computational account of the production effect: still playing twenty questions with nature. *Can. J. Exp. Psychol.* **70**, 154–164 (2016).
56. Jamieson, R. K. & Mewhort, D. J. K. Applying an exemplar model to the artificial-grammar task: string completion and performance on individual items. *Q. J. Exp. Psychol.* **63**, 1014–1039 (2010).
57. Curtis, E. T. & Jamieson, R. K. Computational and empirical simulations of selective memory impairments: converging evidence for a single-system account of memory dissociations. *Q. J. Exp. Psychol.* **72**, 798–817 (2019).
58. Curtis, E. T. Interactive processes in an instance model of memory: a computational analysis of Jacoby's (1983) dissociation between perception and recognition. *Can. J. Exp. Psychol.* **73**, 288–294 (2019).
59. Kwantes, P. J. & Mewhort, D. J. K. Modeling lexical decision and word naming as a retrieval process. *Can. J. Exp. Psychol.* **53**, 306–315 (1999).
60. Clark, S. E. A familiarity-based account of confidence–accuracy inversions in recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **23**, 232–238 (1997).
61. Johns, B. T. & Jones, M. N. Generating structure from experience: a retrieval-based model of language processing. *Can. J. Exp. Psychol.* **69**, 233–251 (2015).

62. Johns, B. T. & Jones, M. N. Perceptual inference through global lexical similarity: topics in cognitive science. *Top. Cognit. Sci.* **4**, 103–120 (2012).

63. Posner, M. I. & Keele, S. W. On the genesis of abstract ideas. *J. Exp. Psychol.* **77**, 353–363 (1968).

64. Posner, M. I. & Keele, S. W. Retention of abstract ideas. *J. Exp. Psychol.* **83**, 304–308 (1970).

65. Zaki, S. R., Nosofsky, R. M., Jessup, N. M. & Unverzagt, F. W. Categorization and recognition performance of a memory-impaired group: evidence for single-system models. *J. Int. Neuropsychol. Soc.* **9**, 394–406 (2003).

66. Deese, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. Exp. Psychol.* **58**, 17–22 (1959).

67. Roediger, H. L. & McDermott, K. B. Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cognit.* **21**, 803–814 (1995).

68. Brainerd, C. J. & Reyna, V. F. Fuzzy-trace theory and false memory. *Curr. Dir. Psychol. Sci.* **11**, 6 (2002).

69. Johns, B. T., Jones, M. N. & Mewhort, D. J. K. A continuous source reinstatement model of true and false recollection. *Can. J. Exp. Psychol.* **75**, 1–18 (2021).

70. Johns, B. T., Jones, M. N. & Mewhort, D. J. K. A synchronization account of false recognition. *Cognit. Psychol.* **65**, 486–518 (2012).

71. Singer, M. in *Learning and Memory: A Comprehensive Reference* (ed. Byrne, J. H.) 357–381 (Elsevier, 2017).

72. Singer, M. & Spear, J. Validation of strongly presupposed text concepts in reading comprehension: cleft constructions. *Can. J. Exp. Psychol.* **74**, 1–11 (2020).

73. Brewer, W. F. & Treyens, J. C. Role of schemata in memory for places. *Cognit. Psychol.* **13**, 207–230 (1981).

74. Knowlton, B. & Squire, L. The learning of categories: parallel brain systems for item memory and category knowledge. *Science* **262**, 1747–1749 (1993).

75. Zaki, S. R. Is categorization performance really intact in amnesia? A meta-analysis. *Psychonomic Bull. Rev.* **11**, 1048–1054 (2004).

76. Gregory, E., McCloskey, M. & Landau, B. Profound loss of general knowledge in retrograde amnesia: evidence from an amnesic artist. *Front. Hum. Neurosci.* **8**, 287 (2014).

77. Gregory, E., McCloskey, M., Ovans, Z. & Landau, B. Declarative memory and skill-related knowledge: evidence from a case study of amnesia and implications for theories of memory. *Cognit. Neuropsychol.* **33**, 220–240 (2016).

78. Renoult, L., Irish, M., Moscovitch, M. & Rugg, M. D. From knowing to remembering: the semantic–episodic distinction. *Trends Cognit. Sci.* **23**, 1041–1057 (2019).

79. Benjamin, A. S., Diaz, M., Matzen, L. E. & Johnson, B. Tests of the DRYAD theory of the age-related deficit in memory for context: not about context, and not about aging. *Psychol. Aging* **27**, 418–428 (2012).

80. Benjamin, A. S. Representational explanations of "process" dissociations in recognition: the DRYAD theory of aging and memory judgments. *Psychol. Rev.* **117**, 1055–1079 (2010).

81. Nosofsky, R. M. & Zaki, S. R. Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation. *Psychol. Sci.* **9**, 247–255 (1998).

82. Nosofsky, R. M., Little, D. R. & James, T. W. Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proc. Natl Acad. Sci. USA* **109**, 333–338 (2012).

83. Zaki, S. R. & Nosofsky, R. M. A single-system interpretation of dissociations between recognition and categorization in a task involving object-like stimuli. *Cognit. Affect. Behav. Neurosci.* **1**, 344–359 (2001).

84. Ashby, F. G. & Rosedahl, L. A neural interpretation of exemplar theory. *Psychol. Rev.* **124**, 472–482 (2017).

85. Malmberg, K. J., Zeelenberg, R. & Shiffrin, R. M. Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by midazolam. *J. Exp. Psychol. Learn. Mem. Cognit.* **30**, 540–549 (2004).

86. Chomsky, N. Rules and representations. *Behav. Brain Sci.* **3**, 1–15 (1980).

87. Ambridge, B. Against stored abstractions: a radical exemplar model of language acquisition. *First Lang.* **40**, 509–559 (2019).

88. Truscott, J. Instance theory and universal grammar in second language research. *Second. Lang. Res.* **14**, 257–291 (1998).

89. Jones, M. N. When does abstraction occur in semantic memory: insights from distributional models. *Lang. Cognit. Neurosci.* **34**, 1338–1346 (2019).

90. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. Preprint at https://arxiv.org/abs/1301.3781 (2013).

91. Lund, K. & Burgess, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Meth. Instrum. Comput.* **28**, 203–208 (1996).

92. Pennington, J., Socher, R. & Manning, C. in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)* (eds Moschitti, A. et al.) 1532–1543 (Association for Computational Linguistics, 2014).

93. Taler, V., Johns, B. T., Young, K., Sheppard, C. & Jones, M. N. A computational analysis of semantic structure in bilingual verbal fluency performance. *J. Mem. Lang.* **69**, 607–618 (2013).

94. Aujla, H. Language experience predicts semantic priming of lexical decision. *Can. J. Exp. Psychol.* **75**, 235–244 (2021).

95. Johns, B. T. & Jamieson, R. K. The influence of place and time on lexical behavior: a distributional analysis. *Behav. Res.* **51**, 2438–2453 (2019).

96. Montag, J. L., Jones, M. N. & Smith, L. B. The words children hear: picture books and the statistics for language learning. *Psychol. Sci.* **26**, 1489–1496 (2015).

97. Griffiths, T. L., Steyvers, M. & Tenenbaum, J. B. Topics in semantic representation. *Psychol. Rev.* **114**, 211–244 (2007).

98. Beekhuizen, B., Armstrong, B. C. & Stevenson, S. Probing lexical ambiguity: word vectors encode number and relatedness of senses. *Cogn. Sci.* **45**, e12943 (2021).

99. Bürki, A. Variation in the speech signal as a window into the cognitive architecture of language production. *Psychon. Bull. Rev.* **25**, 1973–2004 (2018).

100. Disner, S. F. Evaluation of vowel normalization procedures. *J. Acoustical Soc. Am.* **67**, 253–261 (1980).

101. Gerstman, L. Classification of self-normalized vowels. *IEEE Trans. Audio Electroacoust.* **16**, 78–80 (1968).

102. Goldinger, S. D., Pisoni, D. B. & Logan, J. S. On the nature of talker variability effects on recall of spoken word lists. *J. Exp. Psychol. Learn. Mem. Cognit.* **17**, 152–162 (1991).

103. Ryalls, B. O. & Pisoni, D. B. The effect of talker variability on word recognition in preschool children. *Dev. Psychol.* **33**, 441–452 (1997).

104. Thiessen, E. D. & Pavlik, P. I. iMinerva: a mathematical model of distributional statistical learning. *Cogn. Sci.* **37**, 310–343 (2013).

105. Miller, R. R. Challenges facing contemporary associative approaches to acquired behavior. *CCBR* **1**, 77–93 (2006).

106. Stout, S. C. & Miller, R. R. Sometimes-competing retrieval (SOCR): a formalization of the comparator hypothesis. *Psychol. Rev.* **114**, 759–783 (2007).

107. Miller, R. R., Barnet, R. C. & Grahame, N. J. Assessment of the Rescorla–Wagner model. *Psychol. Bull.* **117**, 363–386 (1995).

108. Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. Visual long-term memory has a massive storage capacity for object details. *Proc. Natl Acad. Sci. USA* **105**, 14325–14329 (2008).

109. Fagot, J. & Cook, R. G. Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proc. Natl Acad. Sci. USA* **103**, 17564–17567 (2006).

110. Vaughan, W. & Greene, S. L. Pigeon visual memory capacity. *J. Exp. Psychol. Anim. Behav. Process.* **10**, 256–271 (1984).

111. Voss, J. L. Long-term associative memory capacity in man. *Psychon. Bull. Rev.* **16**, 1076–1081 (2009).

112. Nickerson, R. S. Short-term memory for complex meaningful visual configurations: a demonstration of capacity. *Can. J. Psychol.* **19**, 155–160 (1965).

113. Nickerson, R. S. A note on long-term recognition memory for pictorial material. *Psychon. Sci.* **11**, 58–58 (1968).

114. Shepard, R. N. Recognition memory for words, sentences, and pictures. *J. Verbal Learn. Verbal Behav.* **6**, 156–163 (1967).

115. Standing, L. Learning 10000 pictures. *Q. J. Exp. Psychol.* **25**, 207–222 (1973).

116. Standing, L., Conezio, J. & Haber, R. N. Perception and memory for pictures: single-trial learning of 2500 visual stimuli. *Psychon. Sci.* **19**, 73–74 (1970).

117. Whittlesea, B. W. A. & Williams, L. D. The source of feelings of familiarity: the discrepancy-attribution hypothesis. *J. Exp. Psychol. Learn. Mem. Cognit.* **26**, 547–565 (2000).

118. Kamin, L. in *Punishment and Aversive Behavior* (eds Campbell, B. A. & Church, R. M.) 279–296 (Appleton-Century-Crofts, 1969).

119. Shanks, D. R. Forward and backward blocking in human contingency judgement. *Q. J. Exp. Psychol. Sect. B* **37**, 1–21 (1985).

120. De Houwer, J. & Beckers, T. Higher-order retrospective revaluation in human causal learning. *Q. J. Exp. Psychol. Sect. B* **55**, 137–151 (2002).

121. Matzel, L. D., Schachtman, T. R. & Miller, R. R. Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learn. Motiv.* **16**, 398–412 (1985).

122. Miller, R. R. & Witnauer, J. E. Retrospective revaluation: the phenomenon and its theoretical implications. *Behav. Process.* **123**, 15–25 (2016).

123. Van Hamme, L. J. & Wasserman, E. A. Cue competition in causality judgments: the role of nonpresentation of compound stimulus elements. *Learn. Motiv.* **25**, 127–151 (1994).

124. De Houwer, J., Thomas, S. & Baeyens, F. Association learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychol. Bull.* **127**, 853–869 (2001).

125. Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F. & Crombez, G. Evaluative conditioning in humans: a meta-analysis. *Psychol. Bull.* **136**, 390–421 (2010).

126. Biegler, P. & Vargas, P. Ban the Sunset? Nonpropositional content and regulation of pharmaceutical advertising. *Am. J. Bioeth.* **13**, 3–13 (2013).

127. Brown, G. D. A., Hulme, C. & Preece, T. Oscillator-based memory for serial order. *Psychol. Rev.* **107**, 127–181 (2000).

128. Lipp, O. V., Mallan, K. M., Libera, M. & Tan, M. The effects of verbal instruction on affective and expectancy learning. *Behav. Res. Ther.* **48**, 203–209 (2010).

129. Lipp, O. V., Oughton, N. & LeLievre, J. Evaluative learning in human Pavlovian conditioning: extinct, but still there? *Learn. Motiv.* **34**, 219–239 (2003).

130. Marr, D. *Vision: A Computational Investigation Into The Human Representation and Processing of Visual Information* (ed. Freeman, W. H.) (MIT Press, 1982).

131. Poggio, T. On holographic models of memory. *Kybernetik* **12**, 237–238 (1973).

132. Gabor, D. Associative holographic memories. *IBM J. Res. Dev.* **13**, 156–159 (1969).

133. Franklin, D. R. J. & Mewhort, D. J. K. Memory as a hologram: an analysis of learning and recall. *Can. J. Exp. Psychol.* **69**, 115–135 (2015).

134. Kelly, M. A., Blostein, D. & Mewhort, D. J. K. Encoding structure in holographic reduced representations. *Can. J. Exp. Psychol.* **67**, 79–93 (2013).

135. Kelly, M. A., Mewhort, D. J. K. & West, R. L. The memory tesseract: mathematical equivalence between composite and separate storage memory models. *J. Math. Psychol.* **77**, 142–155 (2017).

136. McClelland, J. L. & Rumelhart, D. E. Distributed memory and the representation of general and specific information. *J. Exp. Psychol.Gen.* **114**, 159–188 (1985).

137. Vokey, J. R. & Higham, P. A. Opposition logic and neural network models in artificial grammar learning. *Conscious. Cognit.* **13**, 565–578 (2004).

138. Vokey, J. R. & Jamieson, R. K. A visual-familiarity account of evidence for orthographic processing in baboons (*Papio papio*). *Psychol. Sci.* **25**, 991–996 (2014).

139. Nosofsky, R. M. & Palmeri, T. J. A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychon. Bull. Rev.* **5**, 345–369 (1998).

140. Nosofsky, R. M. & Palmeri, T. Rule-plus-exception model of classification learning. *Psychol. Rev.* **101**, 53–79 (1994).

141. Erickson, M. A. & Kruschke, J. K. Rules and exemplars in category learning. *J. Exp. Psychol.Gen.* **127**, 107–140 (1998).

142. Brooks, L. R. & Hannah, S. D. Instantiated features and the use of 'rules. *J. Exp. Psychol. Gen.* **135**, 133–151 (2006).

143. Logan, G. D. Automaticity and reading: perspectives from the instance theory of automatization. *Read. Writ. Q.* **13**, 123–146 (1997).

144. Kolers, P. A. Remembering operations. *Mem. Cognit.* **1**, 347–355 (1973).

145. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

146. Simon, H. A. *The Sciences of the Artificial* (MIT Press, 2008).

147. Simon, H. A. Rational choice and the structure of the environment. *Psychol. Rev.* **63**, 129–138 (1956).
148. Gigerenzer, G. & Brighton, H. Homo heuristicus: why biased minds make better inferences. *Top. Cognit. Sci.* **1**, 107–143 (2009).
149. Shiffrin, R. M. Is it reasonable to study decision-making quantitatively? *Top. Cogn. Sci.* https://doi.org/10.1111/tops.12541 (2021).
150. Chater, N. & Oaksford, M. The rational analysis of mind and behavior. *Synthese* **122**, 93–131 (2000).
151. Shi, L., Griffiths, T. L., Feldman, N. H. & Sanborn, A. N. Exemplar models as a mechanism for performing Bayesian inference. *Psychon. Bull. Rev.* **17**, 443–464 (2010).
152. Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* **27**, 125–140 (1962).
153. Shepard, R. Toward a universal law of generalization for psychological science. *Science* **237**, 1317–1323 (1987).
154. Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., Mckinley, S. C. & Glauthier, P. Comparing modes of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961). *Mem. Cogn.* **22**, 352–369 (1994).
155. Sheu, C.-F. A note on the multiple-trace memory model without simulation. *J. Math. Psychol.* **36**, 592–597 (1992).
156. Jamieson, R. K. & Pexman, P. M. Moving beyond 20 questions: we (still) need stronger psychological theory. *Can. Psychol.* **61**, 273–280 (2020).
157. Hirshman, E., Fisher, J., Henthorn, T., Arndt, J. & Passannante, A. Midazolam amnesia and dual-process models of the word-frequency mirror effect. *J. Mem. Lang.* **47**, 499–516 (2002).
158. Arndt, J., Passannante, A. & Hirshman, E. The effect of midazolam on implicit and explicit memory in category exemplar production and category cued recall. *Memory* **12**, 158–173 (2004).
159. Fisher, J., Hirshman, E., Henthorn, T., Arndt, J. & Passannante, A. Midazolam amnesia and short-term/working memory processes. *Conscious. Cognit.* **15**, 54–63 (2006).
160. Hirshman, E., Fisher, J., Henthorn, T., Arndt, J. & Passannante, A. Midazolam amnesia and retrieval from semantic memory: developing methods to test theories of implicit memory. *Brain Cognit.* **53**, 427–432 (2003).
161. Hirshman, E., Passannante, A. & Henzler, A. The effect of midazolam on implicit memory tests. *Brain Cognit.* **41**, 351–364 (1999).
162. Joordens, S. & Hockley, W. E. Recollection and familiarity through the looking glass: when old does not mirror new. *J. Exp. Psychol. Learn. Mem. Cognit.* **26**, 1534–1555 (2000).
163. Pothos, E. M. Theories of artificial grammar learning. *Psychol. Bull.* **133**, 227–244 (2007).
164. Reber, A. S. Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* **6**, 855–863 (1967).
165. Jamieson, R. K. & Mewhort, D. J. K. The influence of grammatical, local, and organizational redundancy on implicit learning: an analysis using information theory. *J. Exp. Psychol. Learn. Mem. Cognit.* **31**, 9–23 (2005).
166. Pothos, E. M. & Bailey, T. M. The role of similarity in artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cognit.* **26**, 847–862 (2000).

Acknowledgements

This work was supported by Natural Sciences and Engineering Research Council of Canada Discovery Grants to R.K.J. and B.T.J. We thank M. J. C. Crump for discussion and suggestions.

Author contributions

R.K.J. was lead author of this manuscript; all other authors contributed equally.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Psychology thanks Charles Brainerd, Gordon Logan and the other, anonymous, reviewer for their contribution to the peer review of this work.

Publisher's note

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