A model for learning structured representations of similarity and relative magnitude from experience

Leonidas AA Doumas[1] and Andrea E Martin[2,3]

How a system represents information tightly constrains the kinds of problems it can solve. Humans routinely solve problems that appear to require abstract representations of stimulus properties and relations. How we acquire such representations has central importance in an account of human cognition. We briefly describe a theory of how a system can learn invariant responses to instances of similarity and relative magnitude, and how structured, relational representations can be learned from initially unstructured inputs. Two operations, comparing distributed representations and learning from the concomitant network dynamics in time, underpin the ability to learn these representations and to respond to invariance in the environment. Comparing analog representations of absolute magnitude produces invariant signals that carry information about similarity and relative magnitude. We describe how a system can then use this information to bootstrap learning structured (i.e., symbolic) concepts of relative magnitude from experience without assuming such representations a priori.

Addresses
[1] Department of Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, United Kingdom
[2] Language and Computation in Neural Systems Group, Max Planck Institute for Psycholinguistics, The Netherlands
[3] Donders Centre for Cognitive Neuroimaging, Radboud University, The Netherlands

Corresponding author: Doumas, Leonidas AA (alex.doumas@ed.ac.uk)

Introduction

Similarity and magnitude computations persist in natural systems. The capacity to appreciate similarities and differences is important across species [1,2]. Knowing that the current food is safe because one has eaten it before requires at least some capacity to appreciate an equivalence between the current perception of the food and a previous one. Similarly, knowing that the current location is novel, and that dead reckoning must be used to find one's way home likely require estimation of magnitude [3].

On the other hand, humans appear to exhibit much more powerful and abstract concepts of similarity and magnitude. For instance, even young children can appreciate that a monotonic increase in size is like a monotonic increase in brightness and use that information to solve problems [4]. We have previously proposed that this ability stems from our capacity to learn structured (i.e., symbolic) representations of relations like 'same', 'different', 'more', and 'less' [5,6,7,8]. Some have gone so far as to claim that the capacity to learn such structured representations and use them to solve problems is the defining aspect of human intelligence that separates us from our non-human cousins [9].

Providing an account of how structured representations of similarity and relative magnitude are learned from experience has proven difficult (e.g., [10–12]). Models that rely on structured representations either make strong nativist claims, positing that a set of structured representational elements and rules for building compositions of these elements are innate, or, at the very least, require that the powerful representations that they use are hand-coded by the modeler (e.g., [13–15]). Requiring assumptions of large banks of a priori representations has been levied as one of the fundamental limitations of the symbolic approach to understanding cognition (e.g., [11,12,16–18]).

At least part of the problem stems from the fact that representations have both content and format. Representations are about something (e.g., standing as an equivalence class of something in the world), specifying information about that thing—their content. That information also has a format. It is organised in some manner—for example, as a list of features, a predicate, or a labeled graph. As such, a complete account of how we learn structured representations of abstract relations from experience entails solving problems relating to both content and format. Doing so entails solving at least three problems: First, the perceptual/cognitive system must detect the basic featural *invariants* that remain constant across instances of the relation. Consequently, the perceptual system must deliver, or learn to deliver, an invariant response to instances of the relation to be learned. Second, the system must isolate those invariants from the other properties of the objects engaged in the relation to be learned. That is, given an activation vector delivered by the perceptual system in response to a

stimulus (e.g., a visual scene wherein a table is larger than a hammer), the system must isolate those properties that code for the relation to be learned (e.g., the invariant features of *larger*). Third, the system must learn a structured, or a *predicate*, representation of the critical invariant properties—that is, it must come to represent those invariant properties as an explicit entity that can be bound to arbitrary and novel arguments while remaining independent of those arguments.

We have previously proposed a theory for how the human cognitive system might learn structured predicate representations from holistic inputs without assuming predicate representations a priori—addressing the second and third of the above problems. The theory is instantiated in a computational model called DORA (Discovery of relations by analogy; [6•]). More recently we have extended the model to address the first problem of invariant discovery [19,20]. This version of the model addresses the question of how a system might learn the content of representations of abstract concepts like 'same', 'more', and 'less'. In brief, the model learns invariant featural (i.e., subsymbolic) responses to instances of similarity, difference, and relative magnitude from absolute encodings of magnitude (e.g., pixel images).

The goal of the present paper is to briefly describe how structured representations of similarity and relative magnitude can be learned from experience. We give a very broad overview of the model, and we focus on the solution to the problem of learning the content of representations of similarity and relative magnitude (SRM). We begin by discussing the problem of learning SRM invariants, and then briefly summari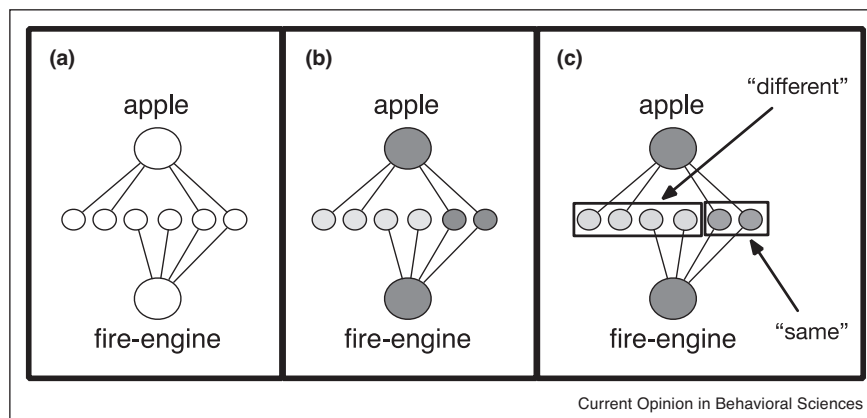se DORA's representation learning algorithm. We then review a few recent results indicating the power of a system that integrates both learning the content and structure of symbolic (predicate-like) representations of SRM, and how it captures important aspects of human cognition.

Learning structured representations of similarity and relative magnitude
Learning content: invariant responses to similarity and difference

Diagnostic and information rich patterns emerge when a system based on distributed representations compares items. Consider a very simple case where the system compares a distributed representation of an apple to a distributed representation of a fire-engine. A node in the network is connected to a set of features coding for the apple in a distributed fashion, and another node in the network is connected to a set of features coding for the fire-engine in a distributed fashion (Figure 1a). When the two objects are compared, the units coding for the apple and those coding for the fire-engine are co-activated (Figure 1b). When both the apple and fire-engine representations are co-active, any features common to both the apple and the fire-engine will receive input from two sources and thus become roughly twice as active as any features unique to either the apple or the fire-engine (Figure 1c). This process of comparison highlights the intersection of the apple and the fire-engine (i.e., their shared features). Comparison-based intersection discovery has been exploited in models like LISA [21•,22] and DORA to support processes such as analogical mapping, schema induction, and learning structured representations from unstructured representations of simple objects.

Figure 1

An illustration of comparison-based intersection highlighting. **(a)** When objects coded by distributed feature sets are compared (i.e., are co-activated), shared features receive input from multiple sources, while features unique to only one of the objects receive input only from a single source. **(b)** Shared features, therefore, become more active than unshared features. **(c)** This difference in activation marks what is similar and different about the compared objects: More active features are properties that are similar, while less active features are properties on which the two objects are different.

Crucially, the pattern of differentially active units that arises during comparison—with similarities marked by more active units and differences by less active units (Figure 1c)—will arise whenever two distributed representations are co-activated. As a consequence, the signal serves as an invariant indicator for basic similarity (or sameness) and difference. Learning an invariant response for sameness or difference, therefore, reduces to learning to respond differentially to these two patterns. As we show below, an invariant feature for 'same' and 'different' learned in this manner supports solving very simple similarity and difference problems, and also serves as the basis for learning structured (i.e., symbolic) representations of abstract *same* and *different* that support more complex problem solving.

This idea is, in its essence, rests on an idea that is similar to that proposed by Wasserman and colleagues (e.g., [2•,23,24•,25]): that entropy underlies same and different detection. Wasserman and colleagues have argued that animals (including humans) are sensitive to entropy present in scenes. In a series of elegant experiments (e.g., [23,24•,25]) they have repeatedly shown that the similarity judgements of animals can be accurately modelled using Shannon entropy. In short, in displays of simple icons, a variety of animals including pigeons, baboons, and humans appear sensitive to the overall entropy in the display (more mismatched icons producing higher entropy), and the probability of categorizing a display as an instance of same or different is a function of the stimulus entropy. Our proposal is that something like the local entropy arising when object properties and dimensions (such as absolute size) are compared or co-activated serves as the basis for learning invariant features for SRM, which can then serve as the content for learning structured representations of these concepts.

Invariant responses to relative magnitude

Just as basic comparison of distributed representations serves to mark basic 'sameness' and 'difference', so comparison also serves as the basis for responding in an invariant manner to relative magnitude information (i.e., 'moreness', 'lessness', and 'sameness').

A comparison-based solution to the problem of learning an invariant feature response for 'more', 'less', and 'same' requires the assumption that initial absolute magnitude information is coded by a direct neural proxy: All else being equal, higher magnitude items are coded by more neurons—or with a higher rate of firing, or signal amplitude, and so on—than comparatively lower magnitude items. There is a large amount of evidence for this assumption. For example, in visual processing, larger items take up more space on the retina, and are coded by larger swaths of the visual cortex (e.g., [26,27]).

Basic magnitude computation follows naturally from this coding mechanism. To illustrate, consider a case where two objects of different size are compared (Figure 2a). If the two representations of absolute size are co-activated, and the units coding the two objects then compete to become active, or to stay active (e.g., via lateral inhibition; Figure 2b), then the larger object will tend to become active first (Figure 2c). When that object becomes inactive (e.g., via a yoked inhibitor unit—see Refs. [6•,21•,22,28•,29]), the smaller object can become active (Figure 2d). In brief, the process of attempting to co-activate two competing representations of objects that code for some form of absolute magnitude information will result in a pattern of firing wherein the item coding the larger magnitude becomes active at a different time scale than the item coding the smaller magnitude. By contrast, when two items of the same magnitude are compared, the two items will settle into a state of mutual co-activation (i.e., both units will remain somewhat active until their inhibitors fire). Moreover, the time to settling serves as a reliable signal for the degree of similarity, with more comparisons of more similar magnitudes taking longer to resolve.
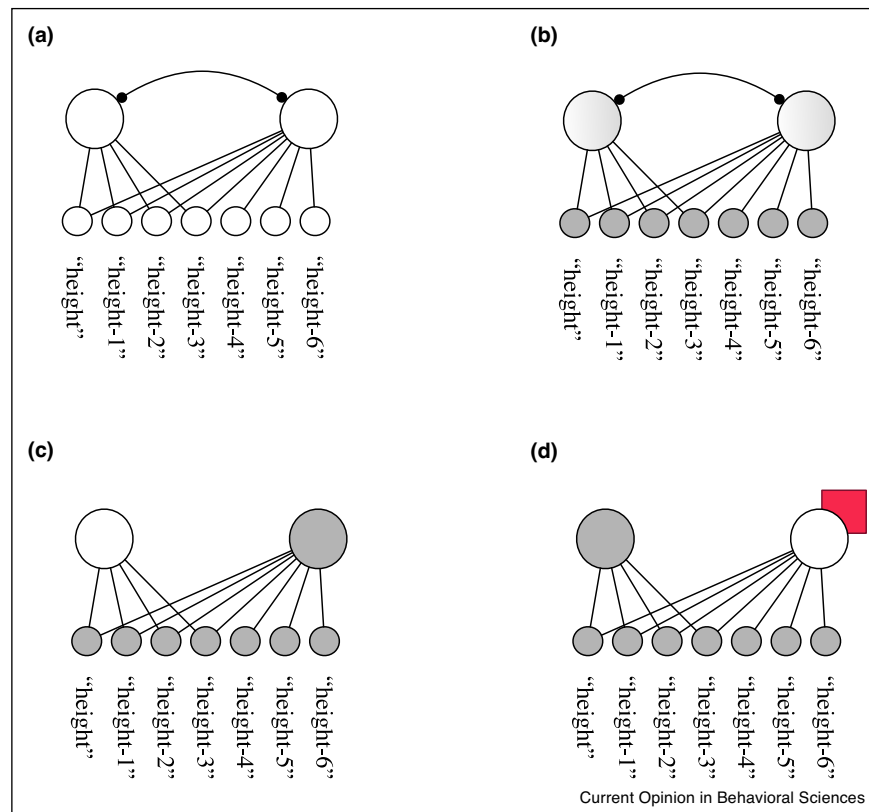
Consequently, comparing items with values on dimensions or extents result in invariant signals for basic relative magnitude information (e.g., basic 'moreness', 'lessness', and 'sameness'). Learning invariant features to code for 'more', 'less', and 'same', therefore, reduces to learning to respond differentially to these patterns. For example, the invariant feature for 'more' are simply be the node (or nodes) that respond to the compared item that becomes active first; the invariant feature for 'less' is then the node (or nodes) that respond to the compared item that that becomes active next; the invariant feature 'same' is the node or (nodes) that respond to two mutually co-active items. These invariant features then form the basis for learning structured (i.e., symbolic) representations of abstract *more*, *less*, and *same* that support more complex problem solving.

Learning format: learning structured representations

As noted above, DORA learns structured (i.e., functionally symbolic) representations from representations of flat feature vectors, and it does so without feedback (i.e., through unsupervised learning; [6•]). In this section we very briefly summarise how it does so. Full details of the model's format learning operation appear in Appendix A of Ref. [6•].

DORA is a settling neural network model, with a long-term memory (LTM) consisting of layered units (called *tokens*) that are connected to a common pool of feature (or perceptual) units. DORA's working memory (WM) consists of two independent sets of units from LTM, the driver—the current focus of attention—and the recipient—or active memory (potentiated units from

Figure 2

An illustration of the basic invariant pattern that emerges when two items coding basic absolute magnitude information compete to become active. **(a)** Two units (large circles) coding for absolute magnitude. **(b)** The semantics (small circles) coding for absolute magnitude are activated, and the two units compete (via lateral inhibition; line with circle endpoint) to become active. **(c)** The unit attached to the greater magnitude wins the competition and becomes active first. **(d)** After the winning unit is inhibited to inactivity by a yoked inhibitor (represented here by the small square behind large circle), the unit attached to the lesser magnitude becomes active.
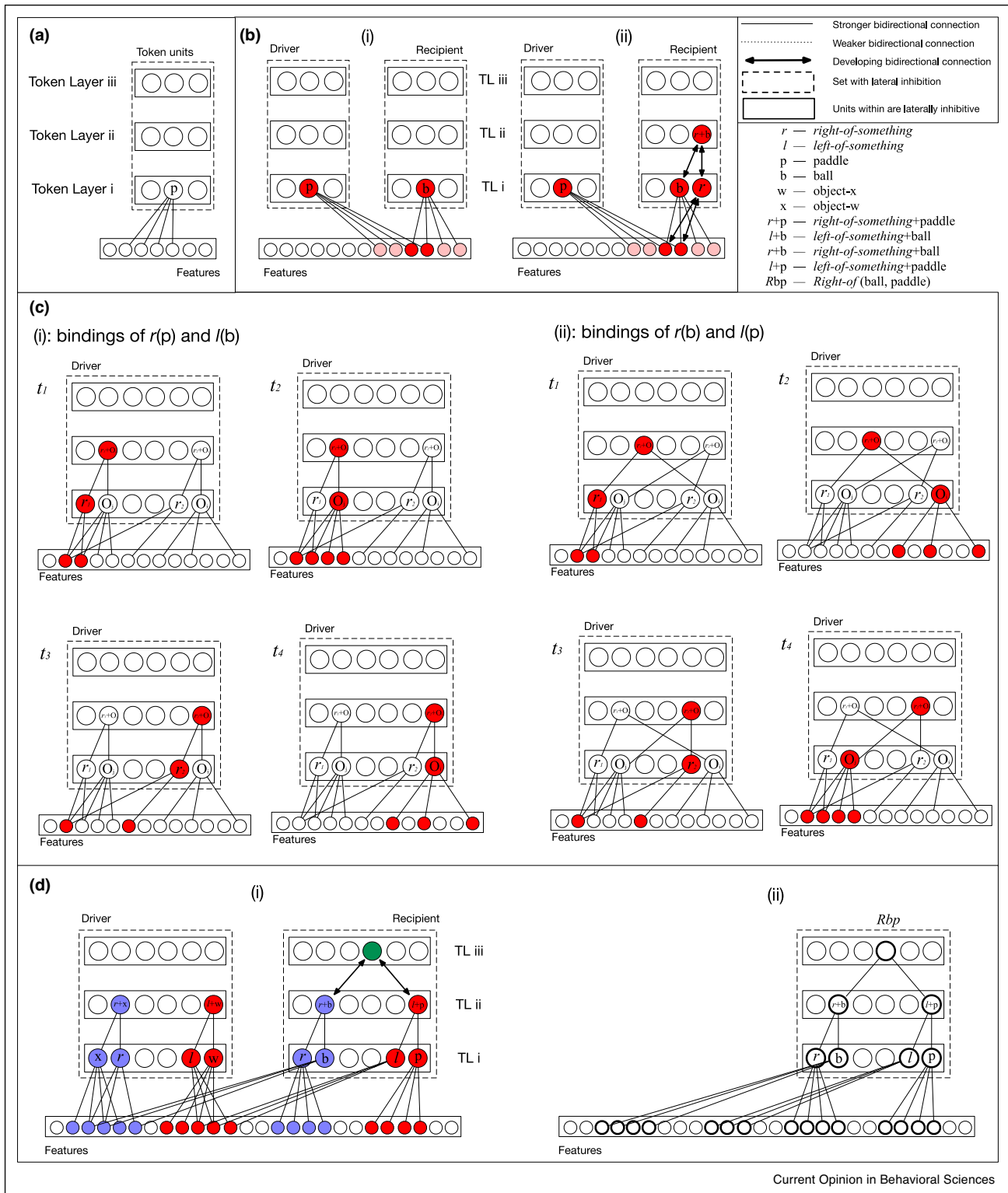
LTM). Activation flows from the driver to the rest of memory via the shared feature units. DORA can learn mapping connections between corresponding coactive units across banks using a modified Hebbian learning algorithm [6[•],20]. Token units are laterally inhibitive within but not across sets.

At a very high level, DORA's learning algorithm has three important features. First, starting from representations of objects encoded as flat feature vectors conjunctively linked by a token unit at token layer I (TLi; Figure 3a), the model isolates and explicitly represents invariant perceptual features or properties (Figure 3b). Comparing (and co-activating) distributed representations of items, O_i and O_j, naturally reveals shared and unshared features of the two items. Specifically, features shared by O_i and O_j will receive roughly twice as much input and, therefore, become roughly twice as active as unshared features (Figure 3bi). DORA exploits this emergent regularity by learning connections between a

recruited unit r_1 in TLi and active feature units, and between a recruited unit in TLii and the active TLi units by Hebbian learning (Figure 3bii). The result is a unit explicitly encoding the intersection of the compared items, and a solution to the isolation problem.

Crucially, these new representations can be bound to arguments via time-based binding (wherein binding information is carried by when units fire), and therefore function as single-place predicates [6[•]]. When laterally inhibitive units are linked by an active conjunctive node, they will naturally oscillate out of synchrony and in direct sequence (phaselag-1). DORA exploits this emergent oscillatory pattern as a binding signal. Specifically, bound units fire in direct sequence, and systematically out of synchrony with other bound pairs (Figure 3ci). In short, identity information is carried by *what* units are active, and binding information is carried by *when* those units are active. This binding signal is dynamic in that the same units are used to represent complementary bindings

Figure 3

Representation learning in DORA. **(a)** DORA starts with representations of objects encoded as flat feature vectors. **(b)** (i) By comparing objects *Oi* and *Oj*, shared features receive more input and become more active. (ii) Using Hebbian learning DORA learns an explicit representation of the featural overlap of b and p (unit labeled *r1*) and links *r1* to Oj (unit labeled *r1*+Oj). **(c)** Illustration of binding in DORA. (i) To bind *r1* to Oj and *r2* to Ok units coding *r1* fire at t1, followed by units for Oj at t2, *r2* at t3 and Ok at t4. (ii) Complementary binding information is carried by a different sequence of firing. **(d)** Learning multi-place relations. (i) Similar predicate-argument sets are compared in driver and recipient. Driver units activate

through different firing orders (Figure 3cii). The resulting representations are functional predicate-object, or role-filler, pairs [6•].

Third, DORA combines lower arity representations into multi-place relational structures (Figure 3d). When DORA compares and aligns two structurally similar sets of bound elements, the aligned (or corresponding) units across the two sets will oscillate systematically (violet and red units in Figure 3di). This pattern only emerges in the model only when role-filler pairs are already linked into a whole relational structure, or when multiple similar cooc-curring role-filler pairs are compared. Thus, the pattern is a reliable signal that DORA exploits to learn multi-place relational structures. As the pattern emerges, DORA learns connections between a unit at a higher layer of the network and active conjunctive nodes by Hebbian learning (Figure 3di). The resulting representation (Figure 3dii) is a functionally symbolic multi-place rela-tional representation. At the layer of feature units, objects, properties, and roles are coded in a distributed fashion. At the next layer, collections of properties are conjunctively coded as tokens of individual objects and predicates. At the next layer, role-filler pairs are conjunc-tively bound. At the top layer, conjunctively coded role-filler pairs are themselves conjunctively coded into whole multi-place relations. The resulting structure allows for long-term storage, and when the representation becomes active, dynamic binding information is carried in the temporal patterns of firing that emerge, with bindings between individual predicates and objects coded by their proximity of firing (as described directly above).

In summary, when DORA compares objects features common to these objects (e.g., their relative value on some dimension) are extracted and represented explicitly on some dimension. These representations are then dynamically bound arguments via time-based binding. Finally, DORA combines sets of predicate-object pairs that co-occur (for example, two instances with one object that is 'more' and another object that is 'less' on a given dimension) to form multi-place structured relational propositions.

DORA's representations in action
DORA's structure learning algorithm has been able to account for over 50 cognitive phenomena, including representation learning and development [6•,30–32], rela-tional learning [33–36], analogy making [6•], and language processing [37–39]. Just like children (see Ref. [40]), DORA starts with the capacity to reason about whole objects, then develops the capacity to reason about iso-lated object properties (e.g., colour), and finally goes through a relational shift (e.g., [41]) and learns structured representations of relational concepts (e.g., *above*, *chases*; see Ref. [7], for a review). Finally, DORA is descended from the LISA model of analogical reasoning proposed by Hummel and Holyoak [21•,22]. DORA's predicate learn-ing algorithm provides an account of how structured representations like those used by LISA can be learned in the first place. As such, DORA basically 'grows up' into LISA as it learns structured relational representations, and the resulting LISA/DORA architecture account for many additional phenomena from the literature on human relational thinking (e.g., [21•,22,42–44]).

When coupled with the capacity to learn to detect simple visual invariants of SRM (i.e., to learn the content of relational representations), DORA learns structured rela-tional representations from simple visual (pixel) images [19]. Recently we have used DORA to learn to play simple video games such as Breakout using a simple Q-learning algorithm to learn policies based on the relations DORA learns from the game or even from unrelated images (e.g., images of shapes on tables). Unlike current machine-learning systems such as deep neural networks, but just like humans, DORA's repre-sentations support cross-domain generalisation. For example, DORA can use representations it has learned playing one video game to play another game without any additional training (i.e., zero-shot learning), to engage in a completely unrelated task such as analogy making, and to solve problems from the developmental literature on category learning [20]. The reason is that unlike systems such as deep neural networks, the repre-sentations that DORA learns are structured, and thus can be freely bound and generalised to novel objects in novel situations (see e.g., [45,46]).

However, while the LISA/DORA framework provides a strong account of a number of phenomena, it also leaves many questions unanswered. One of the interesting lim-itations of the model is in the relation of the representa-tions that it learns and the production of natural language. We have had some limited success modelling aspects of language with DORA (e.g., [37]), and the representations that the model learns on the surface would seem sufficient for modelling some important parts of natural language. For example, LISA/DORA's relational representations are higher-order (i.e., relations can take other relations as arguments), and thus would potentially support hier-archical syntax. However, the representations that DORA

(Figure 3 Legend Continued) featurally similar units in recipient: As violet units in the driver become active they will activate violet units in recipient, and red units in the driver will cause red recipient units to become active. Using Hebbian learning DORA learns a conjunctive encoding of the predicate-argument pairs in recipient as they become active (green unit in Tiii). (ii) The resulting representation encodes a multi-place relational proposition.

learns (and LISA uses) have a particular formal property: specifically, they are role-filler binding representations.

In a role-filler binding system, multi-place predicates are represented as sets of constituent single-place predicates linked via an operator. For example, a relational instance like 'the dog chases the cat' would be represented as something like, *chases* (dog, cat), in a traditional propositional notation. By contrast, in a role-filler binding system the chases relation is represented by single-place predicates specifying its roles linked into a single structure, such as *chaser*(dog) + *chased* (cat). Role-filler binding representations have many properties that make them good models of human representations (see, e.g., [5,47] for reviews). For example, like human mental representations, role-filler binding representations make role information explicit (e.g., [48,49]). However, most natural languages do not follow role-filler binding conventions (although some can, e.g., [50]). Accounting for the interface between the underlying representations posited by LISA/DORA and natural language remains an interesting open question.

Discussion

Humans and many non-human animals appear to have the capacity to detect and respond to SRM (e.g., [2•,24•]). However, the capacities of humans to reason about SRM far outstrip other animals (for a review see Ref. [9•]). A reasonable proposal for the observed disconnect between the relative abilities of humans and non-humans is that humans represent the world differently than non-humans: Specifically, human mental representations are structured (or symbolic; e.g., [45,51]).

As noted above, representations have content and format, and a complete account of representation learning must explain how both of these develop. We have proposed a model, DORA, that addresses both content and format learning. On the one hand, DORA exploits invariant patterns that emerge when absolute magnitude encodings are overlaid. The model learns a simple circuit that responds to these invariant patterns, and these responses become the invariant features properties like 'same', 'different', 'more', and 'less'. On the other hand, DORA learns structured representations from experience without assuming them a priori. By mapping simple objects, DORA isolates invariant features and represents them explicitly. These representations function like single-place predicates as they are dynamically bound to arguments via time-base binding. Finally, via further mappings, co-occurring sets of predicate-object bindings (e.g., 'more' and 'less') are linked to form functional multi-place propositions.

There are a few interesting aspects of the model that bear on the topic of learning representations of SRM. First, all of both DORA's content and format learning, rely

exquisitely on the process of comparison. Comparing two encodings of absolute magnitude produces the invariant patterns that the model exploits to learn invariant featural responses for 'same', 'more', and 'less'. Similarly, comparing two objects serves to highlight their shared properties and bootstraps learning explicit representations of these properties. Finally, comparison serves as a reliable mechanism for discovering similar sets of co-occurring predicate-argument pairs and linking them to form multi-place relational propositions. DORA thus makes an important claim about what kinds of mental operations should support successful learning of SRM concepts, one that has found support in the literature (e.g., [52,53]).

Second, the distinction that the model makes in terms of learning representational content and format provides some interesting possibilities about the differences in the representational capacities of humans and non-humans. The capacity to learn to compare absolute encodings of magnitude may indeed be common to a large swath of animals. Comparing absolute encodings of magnitude provides information about the degree of similarity (e.g., in time to settle, see above), and provides an account for cross-species observations of similarity slopes (e.g., [24•]). By contrast, the capacity to learn an invariant response to SRM signals (e.g., activating a feature for instances of 'same') is sufficient to solve problems like the relational match-to-sample task (see Ref. [54] for a review), and may be less common across species. The capacity to learn structured representations of this content may be what differentiates humans from non-human species and gives rise to the uniquely human reasoning capacities discussed in, for instance, Ref. [9•].

On a similar note, a consistent trend that has emerged in developmental science is the relational shift, or the emergent capacity of children to solve analogical problems starting around the age of five [41,55]. However, recent work (e.g., [56,57]) has shown that even very young infants reason about some simple relational concepts like same. The separation of relational format and content learning that DORA makes explicit provides a potential reconciliation of these results. We and others have posited that the observed relational shift in humans might be a product of learning a specific type of representation (see Ref. [6•]). Specifically, the capacity of children to solve analogy type problems after age five or so might be a result of learning structured representations that support explicit analogical mapping [41]. However, the capacity to detect relational information and to use it to solve some kinds of problems might be present much earlier.

Third, we emphasize the fact that the theory and model presented here is able to discover and learn structured representations from experience because it exploits regularities of network behavior in a distributed computing

system. We believe that this solution represents a fundamental formal and neurophysiological alignment between how human-like representations can be achieved in a system that learns, and how distributed neural computing systems, including neural assemblies, process information (see Ref. [29] for a historical precedent). Neural oscillations, or brain rhythms [58] have long been implicated as indices of neural information processing (e.g., [59[•]]) in many systems, even when those systems differ in scale and configuration [60] and have changed the nature of theories of the neural implementation of human memory (e.g., [61]) and speech and language processing (e.g., [37,38,62]).

Conflict of interest statement
Nothing declared.

Acknowledgements

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest

1. Premack D: **Animal cognition**. *Ann Rev Psychol* 1983, **34**:351-362.

2. Wasserman EA, Young ME: **Same–different discrimination: the**
• **keel and backbone of thought and reasoning**. *J Exp Psychol* 2010, **36**:3-22
A review of the argument that same and difference detection is a fundamental component of animal cognition and a statement of the role of entropy on said processing.

3. Gallistel CR: *The Organization of Learning*. The MIT Press; 1990
A classic statement of the problem of learning representations that support structured thinking.

4. Kotovsky L, Gentner D: **Comparison and categorization in the development of relational similarity**. *Child Dev* 1996, **67**:2797-2822.

5. Doumas LAA, Hummel JE: **Computational models of higher cognition**. *The Oxford Handbook of Thinking and Reasoning*. New York, NY: Oxford University Press; 2012, 52-66.

6. Doumas LAA, Hummel JE, Sandhofer CM: **A theory of the**
• **discovery and predication of relational concepts**. *Psychol Rev* 2008, **115**:1-43
The full model description of the architectures, algorithms, and principles needed to learn relational predicates from flat feature vectors. Also features simulations of human data from the literature on developmental relational reasoning.

7. Doumas LAA, Martin AE: **Learning structured representations from experience**. *Psychology of Learning and Motivation*. Academic Press; 2018:165-203.

8. Martin AE, Doumas LAA: **Predicate learning in neural systems: using oscillations to discover latent structure**. *Curr Opin Behav Sci* 2019, **29**:77-83.

9. Penn DC, Holyoak KJ, Povinelli DJ: **Darwin's mistake: explaining**
• **the discontinuity between human and nonhuman minds**. *Behav Brain Sci* 2008, **31**:109-130
A very comprehensive overview of the argument that humans alone appear to exhibit truly relational cognition.

10. Kriete T, Noelle DC, Cohen JD, O'Reilly RC: **Indirection and symbol-like processing in the prefrontal cortex and basal ganglia**. *Proc Natl Acad Sci U S A* 2013, **110**:16390-16395.

11. Leech R, Mareschal D, Cooper RP: **Analogy as relational priming: a developmental and computational perspective on the origins of a complex cognitive skill**. *Behav Brain Sci* 2008, **31**:357-378.

12. Rogers TT, McClelland JL: **Précis of semantic cognition: a parallel distributed processing approach**. *Behav Brain Sci* 2008, **31**:689.

13. Goodman ND, Ullman TD, Tenenbaum JB: **Learning a theory of causality**. *Psychol Rev* 2011, **118**:110.

14. Kemp C, Tenenbaum JB: **Structured statistical models of inductive reasoning**. *Psychol Rev* 2009, **116**:20.

15. Lake BM, Salakhutdinov R, Tenenbaum JB: **Human-level concept learning through probabilistic program induction**. *Science* 2015, **350**:1332-1338.

16. O'Reilly RC, Busby RS: **Generalizable relational binding from coarse-coded distributed representations**. *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press; 2002, 75-82.

17. O'Reilly RC, Busby RS, Soto R: **Three forms of binding and their neural substrates: alternatives to temporal synchrony**. *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford, England: Oxford University Press; 2003, 168-192.

18. Rumelhart DE, McClelland JL, the PDP Research Group (Eds): *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol 1. Cambridge, MA: MIT Press; 1986.

19. Doumas LAA, Hamer A, Puebla G, Martin AE: **A theory of the detection and learning of structured representations of similarity and relative magnitude**. *The 39th Annual Conference of the Cognitive Science Society (CogSci 2017)*; *Cognitive Science Society: 2017:1955-1960*.

20. Doumas LAA, Puebla G, Martin AE, Hummel JE: **Relation learning in a neurocomputational architecture supports cross-domain transfer**. *Annual Conference of the Cognitive Science Society (CogSci 2020)*; *Cognitive Science Society: 2020:932-937*.

21. Hummel JE, Holyoak KJ: **Distributed representations of**
• **structure: a theory of analogical access and mapping**. *Psychol Rev* 1997, **104**:427
The first paper to demonstrate the success of time in carrying binding information in a model that performs analogical and relational reasoning.

22. Hummel JE, Holyoak KJ: **A symbolic-connectionist theory of relational inference and generalization**. *Psychol Rev* 2003, **110**:220.

23. Wasserman EA, Fagot J, Young ME: **Same–different conceptualization by baboons (Papio papio): the role of entropy**. *J Comp Psychol* 2001, **115**:42.

24. Young ME, Wasserman EA: **Entropy detection by pigeons:**
• **response to mixed visual displays after same–different discrimination training**. *J Exp Psychol* 1997, **23**:157
A clear early statement of the idea that entropy calculation might underlie same/different perception in non-human (and even human) animals.

25. Young ME, Wasserman EA: **Entropy and variability discrimination**. *J Exp Psychol* 2001, **27**:278.

26. Engel SA, Rumelhart DE, Wan dell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN: **fMRI of human visual cortex**. *Nature* 1994, **369**:525.

27. Wandell BA, Dumoulin SO, Brewer AA: **Visual field maps in human cortex**. *Neuron* 2007, **56**:366-383.

28. von der Malsburg C: *The Correlation Theory of Brain Function*
• *(Internal Rep. No. 81–82)*. Goettingen, Germany: Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry; 1981
The first instance that we are aware of a model that uses time to solve the binding problem.

29. von der Malsburg C: **The what and why of binding: the modeler's perspective**. *Neuron* 1999, **24**:95-104.

30. Doumas LAA, Hummel JE: **A computational account of the development of the generalization of shape information**. *Cogn Sci* 2010, **34**:698-712.

31. Doumas LAA, Hummel JE: **Comparison and mapping facilitate relation discovery and predication**. *PLoS One* 2013, **8**:e63889.

32. Morrison RG, Doumas LAA, Richland LE: **A computational account of children's analogical reasoning: balancing inhibitory control in working memory and relational representation**. *Dev Sci* 2011, **14**:516-529.

33. Livins KA, Doumas LAA: **Recognising relations: what can be learned from considering complexity**. *Think Reason* 2015, **21**:251-264.

34. Livins KA, Doumas LAA, Spivey MJ: **Shaping relations: exploiting relational features for visuospatial priming**. *J Exp Psychol* 2016, **42**:127.

35. Livins KA, Spivey MJ, Doumas LAA: **Varying variation: the effects of within-versus across-feature differences on relational category learning**. *Front Psychol* 2015, **6**:129.

36. Son JY, Doumas LAA, Goldstone RL: **When do words promote analogical transfer?** *J Probl Solving* 2010, **3**:4.

37. Martin AE, Doumas LAA: **A mechanism for the cortical computation of hierarchical linguistic structure**. *PLoS Biol* 2017, **15**:e2000663.

38. Martin AE, Doumas LAA: **Tensors and compositionality in neural systems**. *Philos Trans R Soc B* 2020, **375**:20190306.

39. Rabagliati H, Doumas LAA, Bemis DK: **Representing composed meanings through temporal binding**. *Cognition* 2017, **162**:61-72.

40. Smith LB: **Young children's understanding of attributes and dimensions: a comparison of conceptual and linguistic measures**. *Child Dev* 1984:363-380.

41. Gentner D: **Why we're so smart**. *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press; 2003, 195-235.

42. Krawczyk DC, Holyoak KJ, Hummel JE: **Structural constraints and object similarity in analogical mapping and inference**. *Think Reason* 2004, **10**:85-104.

43. Morrison RG, Krawczyk DC, Holyoak KJ, Hummel JE, Chow TW, Miller BL, Knowlton BJ: **A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration**. *J Cogn Neurosci* 2004, **16**:260-271.

44. Viskontas IV, Morrison RG, Holyoak KJ, Hummel JE, Knowlton BJ: **Relational integration, inhibition, and analogical reasoning in older adults**. *Psychol Aging* 2004, **19**:581.

45. Holyoak KJ, Hummel JE: **The proper treatment of symbols in a connectionist architecture**. *Cognitive Dynamics: Conceptual Change in Humans and Machines*. LEA; 2000:229-263.

46. Hummel JE: **Getting symbols out of a neural architecture**. *Connect Sci* 2011, **23**:109-118.

47. Doumas LAA, Hummel JE: **Approaches to modeling human mental representations: what works, what doesn't and why**. *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press; 2005:73-94.

48. Bassok M, Olseth KL: **Object-based representations: transfer between cases of continuous and discrete models of change**. *J Exp Psychol* 1995, **21**:1522.

49. Krawczyk DC, Holyoak KJ, Hummel JE: **The one-to-one constraint in analogical mapping and inference**. *Cogn Sci* 2005, **29**:797-806.

50. Senghas A, Newport EL, Supalla T: **Argument structure in Nicaraguan sign language: the emergence of grammatical devices**. *Proceedings of the 21st Annual Boston University Conference on Language Development* 1997.

51. Markman AB: *Knowledge Representation*. Mahwah, NJ: Lawrence Erlbaum Associates; 1999.

52. Christie S, Gentner D: **Where hypotheses come from: learning new relations by structural alignment**. *J Cogn Dev* 2010, **11**:356-373.

53. Goldwater MB, Don HJ, Krusche MJ, Livesey EJ: **Relational discovery in category learning**. *J Exp Psychol* 2018, **147**:1.

54. Holyoak KJ, Thagard P: *Mental Leaps*. MIT Press; 1995.

55. Richland LE, Morrison RG, Holyoak KJ: **Children's development of analogical reasoning: insights from scene analogy problems**. *J Exp Child Psychol* 2006, **94**:249-273.

56. Hochmann JR, Carey S, Mehler J: **Infants learn a rule predicated on the relation same but fail to simultaneously learn a rule predicated on the relation different**. *Cognition* 2018, **177**:49-57.

57. Ferry AL, Hespos SJ, Gentner D: **Prelinguistic relational concepts: investigating analogical processing in infants**. *Child Dev* 2015, **86**:1386-1405.

58. Buzsaki G: *Rhythms of the Brain*. Oxford University Press; 2006.

59. Singer W: **Synchronization of cortical activity and its putative**
•   **role in information processing and learning**. *Ann Rev Physiol* 1993, **55**:349-374
A review of early work showing evidence for and statement of the idea that the brain uses time to carry information.

60. Buzsáki G: *The Brain from Inside Out*. Oxford University Press; 2019.

61. Lisman JE, Jensen O: **The theta-gamma neural code**. *Neuron* 2013, **77**:1002-1016.

62. Giraud AL, Poeppel D: **Cortical oscillations and speech processing: emerging computational principles and operations**. *Nat Neurosci* 2012, **15**:511.