Behavioral and Brain Sciences (forthcoming)

This Target Article has been accepted for publication and has not yet been copyedited and proofread. The article may be cited using its doi (About doi), but it must be made clear that it is not the final version.

Word counts:

Abstracts: 274

Main text: 14,029

References: 7,107

Entire text: 21,637

The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis Across the Cognitive Sciences[1]

Jake Quilty-Dunn, Washington University in St. Louis, USA, quiltydunn@gmail.com, sites.google.com/site/jakequiltydunn/

Nicolas Porot, Africa Institute for Research in Economics and Social Sciences, Mohammed VI Polytechnic University, Morocco, nicolasporot@gmail.com, nicolasporot.com

Eric Mandelbaum, The Graduate Center & Baruch College, CUNY, USA, eric.mandelbaum@gmail.com, ericmandelbaum.com

Short Abstract: This paper provides a survey of evidence from computational cognitive psychology, perceptual psychology, developmental psychology, comparative psychology, and social psychology, in favor of the language of thought hypothesis (LoTH). We outline six core properties of LoTs and argue that these properties cluster together throughout cognitive science. Instead of regarding LoT as a relic of the previous century, researchers in cognitive science and philosophy of mind should take seriously the explanatory breadth of LoT-based architectures as computational/representational approaches to the mind continue to advance.

[1] All authors contributed equally; authorship is in reverse alphabetical order.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Long Abstract: Mental representations remain the central posits of psychology after many decades of scrutiny. However, there is no consensus about the representational format(s) of biological cognition. This paper provides a survey of evidence from computational cognitive psychology, perceptual psychology, developmental psychology, comparative psychology, and social psychology, and concludes that one type of format that routinely crops up is the language of thought (LoT). We outline six core properties of LoTs: (i) discrete constituents; (ii) role-filler independence; (iii) predicate-argument structure; (iv) logical operators; (v) inferential promiscuity; and (vi) abstract content. These properties cluster together throughout cognitive science. Bayesian computational modeling, compositional features of object perception, complex infant and animal reasoning, and automatic, intuitive cognition in adults all implicate LoT-like structures. Instead of regarding LoT as a relic of the previous century, researchers in cognitive science and philosophy of mind must take seriously the explanatory breadth of LoT-based architectures. We grant that the mind may harbor many formats and architectures, including iconic and associative structures as well as deep-neural-network-like architectures. However, as computational/representational approaches to the mind continue to advance, classical compositional symbolic structures—i.e., LoTs—only prove more flexible and well-supported over time.

Keywords: animal cognition, automaticity, cognitive architecture, deep learning, dual-process theories, implicit attitudes, infant cognition, language of thought, object files, visual cognition

1. Introduction

Mental representations remain the central posits of psychology after many decades of scrutiny. But what are mental representations and what forms do they take in nature? In other words, what is the format of thought? This paper revisits an old answer to this question: The Language of Thought Hypothesis (LoTH).

LoTH is liable to evoke memories of the previous century: foundational discussions about the structure of thought in the 1970s, the rise of connectionism in the 1980s, debates about systematicity and productivity in the 1990s. Now, well into the 21st century, it might seem that LoTH is a relic, like Freud's tripartite cognitive architecture or Skinnerian behaviorism—a topic of historical interest, but no longer at the center of scientific or philosophical inquiry into the mind.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

We will argue for the opposite view: in the half-century since Fodor's (1975) foundational discussion, the case for the LoTH has only grown stronger over time. The chief aim of this paper is to showcase LoTH's explanatory breadth and power in light of recent developments in cognitive science. Computational cognitive science, comparative and developmental psychology, social psychology, and perceptual psychology have all advanced independently, yet evidence from these disparate fields points to the same overall picture: contemporary cognitive science presupposes the language of thought (LoT).

The theoretical literature on LoTH is massive and extremely important for understanding the hypothesis and its historical roots. Given space constraints, we will have to ignore huge portions of this literature. We aim simply to provide the strongest article-sized empirical case for LoTH. As a result, we're forced to ignore a great deal of empirical evidence in favor of LoTH. Work in syntax, semantics, psycholinguistics, and philosophy of mind has often been taken to bolster LoTH (Fodor 1975; 1987). While the relevance of linguistics (broadly construed) to LoTH remains strong, we situate largely independent forms of evidence at the center of our case. We focus primarily on areas (e.g., perception, System-1 reasoning, animal cognition) that seem less language-like. If even these apparent problem areas offer evidence for LoTH, then we should be optimistic about finding evidence for LoTH throughout much of the mind.

In §2, we specify which systems of representation count as LoTs. Some of the conclusions of this section will be a bit surprising, as the natural inferences one should draw from the standard characterization of LoTH have largely been ignored since the view's inception. Then, in §3, §4, §5, and §6, we marshall evidence for LoTH from across the cognitive sciences. §3 reviews recent LoT-based developments in computational cognitive science, §4 surveys a mass of data from the study of human perception, §5 considers evidence from developmental and comparative psychology, and §6 examines evidence from social psychology.

We think that LoTH is indispensable to a computational account of the mind. But the empirical case for the view does not stem from the idea that LoTH is the "only game in town," which it is not (and never really was). Instead, we contend, LoTH is the *best* game in town. For a wide variety of phenomena, it does the best job of explaining why biological minds work in the peculiar ways they do.

Our defense of LoTH doesn't presuppose a single, large-scale opponent. Broadly speaking, our opponents are reductionists of various stripes, e.g., traditional neural reductionists (Churchland 1981; Bickle 2003), theorists who reduce LoT-like cognition to natural language (Berwick &

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Chomsky 2016; Hinzen & Sheehan 2013), critics of representationalism (Hutto & Myin 2013; Schwitzgebel 2013), associationists (Papineau 2003; Rydell & McConnel 2006; Dickinson 2012), and most prominently in recent years, reductionist deep-learning approaches (LeCun, Bengio, & Hinton 2015).[2] However, with the exception of deep neural nets (DNNs), we will mostly avoid direct engagement with these views—not because they are not of interest, but because the best counter to reductionism is simply to demonstrate the explanatory successes of LoT-like representational structures. In the context of System 1 cognition, for example, our primary opponents will be associationist; in the context of perception science, where associationism is less prominent, our foil will be rival iconic/imagistic formats. This focus on multiple corners of cognitive science will demonstrate two rare virtues of LoTH: its unificatory power across disciplines and its generalizability across content domains.

2. What Is a Language of Thought?

Classic defenses of LoTH often equated it with the view that mental representations are *structured* (Fodor 1987; Fodor & Pylyshyn 1988). The route from this identification to the "Only Game in Town" argument is simple—mental representations must have some sort of structure for computational explanations to succeed, and if LoTH follows from that simple fact, it's hard to envision viable alternatives. Arguably, this emphasis on structure *per se* was influenced by the idea that the primary alternatives to LoTH were connectionist models that lacked structured representations altogether (Rumelhart & McClelland 1986; cf. Smolensky 1990).

However, we don't assume this dialectic here. The main reason is that we think there are structured (i.e., non-atomic) representations couched in non-LoT-like formats. Iconic representations are perhaps the clearest example. Operations like mental rotation (Shepard & Metzler 1971) and scanning (Kosslyn, Ball, & Reiser 1978) are inexplicable without appeal to structured representations, but at least some of those representations seem to have an iconic, rather than LoT-like, representational format (Kosslyn 1980; Fodor 2007; Carey 2009; Toribio 2011; Quilty-Dunn 2020b; cf. Pylyshyn 2002). Other potential formats include analog magnitudes (Meck and Church 1983; Carey 2009; Mandelbaum 2013; Clarke 2019; Beck & Clarke forthcoming), vectors in multi-dimensional similarity spaces (Gauker 2011), mental maps

[2] We focus on reductionists because one can grant that, e.g., associative processing and natural-language-guided cognition exist, while also positing a LoT. Our opponents are not theorists who merely posit these mechanisms (as we do), but rather theorists who think all *prima facie* LoT-like cognition reduces to them. See, e.g., Lecun et al.'s argument that the success of DNNs "raises serious doubts about whether understanding a sentence requires anything like the internal symbolic expressions that are manipulated by using inference rules" (2015, 441).

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

(Tolman 1948; Camp 2007; Rescorla 2009; Shea 2018), mental models (Johnson-Laird 2006), graphical models (Danks 2014), semantic pointers (Eliasmith 2013), pattern-separated representations (Yassa & Stark 2011; cf. Quiroga 2020), neural representations at various scales (Barack & Krakauer 2021), and much else. We're happy to let a thousand representational formats bloom.

We take LoTH to describe a representational format with six distinctive properties beyond merely having structure. Many, perhaps all, of these properties are not necessary for a representational scheme to count as a LoT, and some may be shared with other formats. We regard these properties as (somewhat) independent axes on which a format can be assessed for how LoT-like it is. If LoT is a natural kind, then these properties should cluster together homeostatically—i.e., if some properties are instantiated, it raises the probability that others are as well (Boyd 1999). These six features each expand the expressive power of abstract, domain-general cognition, making it advantageous for them to evolve as a cluster. We also suspect there might be distinct LoTs with only partially overlapping properties, perhaps arising in different species or different systems within the same mind. The properties adumbrated here don't necessarily exhaust the characterization of LoTH. The crux of the paper includes several sections devoted to empirical evidence, and a fuller picture of LoTH will emerge throughout.

Before moving to the list of core LoT properties, some caveats about how our approach differs from classic defenses of LoTH. First, while LoTH is sometimes understood as the hypothesis that mental representations have the same structure as natural language, this is not our strategy. While some theorists have posited LoT to explain natural language processing and even play a constitutive role in the compositional semantics of natural language (Fodor 1987; Pinker 1994), our plan is to search for LoT outside natural-language-guided contexts. We will examine LoT-like structures that are less connected to natural language and thus represent stringent test cases for LoTH: mid-level vision, nonverbal minds, and System-1 cognition. LoTH as we'll defend it is committed to representational formats that are language-like in some broad respects, but independent characterizations are provided by both the logical character of LoT (i.e., the way it resembles formal languages that may be radically unlike natural language) and the previous theoretical literature on LoTH, which commits to certain distinctive features. As long as one agrees that an important class of mental representations has many or all of these features, there is no need to quibble about the analogy to natural language.

Second, we will avoid direct discussion of two features of thought that have dominated earlier discussions, namely, systematicity and productivity (Fodor & Pylyshyn 1988). We agree with the widespread view that any format worth calling a LoT must not only have structure, it must be

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

compositional: it must include complex representations that are a function of simple elements plus their mode of combination (cf. Szabo 2011). But as Camp (2007) and others argue, this feature is arguably present in various representational forms, including maps, and thus is not sufficient for ensuring a LoT. Compositionality that is fully systematic and productive is very good *evidence* for LoT-like architectures, but we want to leave open whether some of the LoT-like structures we'll explore are fully systematic and productive. As a historical note, this caveat is in keeping with earlier discussions, in which systematicity and productivity were each considered "a contingent feature of thought" (Fodor 1987, 152) that evidences LoTH rather than a constitutive requirement. This caveat also dovetails with the previous one about relaxing the analogy with natural language—while (e.g.) recursive productivity might be a key feature of natural language (Chomsky 2017), we allow that some LoT-based systems may fail to be recursive. Finally, while we believe systematicity and productivity were good arguments for LoTH, the nature of these cognitive features and their presence in biological minds, including nonverbal ones, is well-trodden ground (Carruthers 2009; Camp 2009). Since our goal is to point in new directions for LoTH, we will invoke systematicity and productivity sparingly, mostly keeping instead to the six core properties listed below. These properties are intended to capture the spirit of earlier presentations of LoTH—a combinatorial, symbolic representational format that facilitates logical, structure-sensitive operations (Fodor & Pylyshyn 1988)—while framing an updated discussion more closely tied to contemporary experimental research.

Property 1: *Discrete constituents*. Typical iconic representations holistically encode features and individuals (Kosslyn, Thompson, & Ganis 2006; Fodor 2007; Hummel 2013), while LoT representations comprise distinct constituents corresponding to individuals and their separable features. In a sentence like "That is a pink square object", the predicate "square" can be deleted without any other constituents being deleted. In an iconic representation of a pink square, the relationship between the individual, its color, and its shape is more intertwined. "Pink square" can be the output of a Merge operation (Chomsky 1995) while the part of the icon that represents pink and the part that represents square are one and the same.

Property 2: *Role-filler independence*. LoT architectures have a distinctive syntax: they combine constituents in a way that maintains independence between syntactic roles and the constituents that fill them (Hummel 2011; Martin & Doumas 2020; Frankland & Greene 2020). The role *agent* is present in "John loves Mary" and "Mary loves John". The identity of the role is independent of what fills it ("Mary", "John"). Likewise, each constituent maintains its identity independent of its current role ("John" can be agent or patient). Role-filler independence captures the rule-based syntactic characteristics of LoT-like compositionality: the syntactic structure is typed independently of its particular constituents, and the constituents are typed independently of

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

how they happen to compose on a particular occasion. In map-like representations, for example, changing the spatial position of a marker changes not only the tputative "predicate" (e.g., *tree*) but also the spatial content of the marker (e.g., its position relative to other map-elements); thus maps fail to exhibit full role-filler independence (Kulvicki 2015). Similarly, connectionist models that bind contents through tensor products (Smolensky 1990; Eliasmith 2013; Palangi, Smolensky, He, & Deng 2018) can simulate compositionality, but fail to preserve identity of the original representational elements; thus they sacrifice role-filler independence, and with it classical compositionality (Hummel 2011; Eliasmith 2013, 125ff).

Role-filler independence might seem similar to the property of having discrete constituents, but they're not equivalent. One could posit discrete constituents in an unordered set, for example, without positing a role that maintains its identity across multiple fillers. There's also nothing in the positing of discrete constituents *per se* that precludes the type-identity of those constituents from shifting in various contexts (e.g., GREEN APPLE and GREEN PEN might be complexes of discrete constituents, but the co-presence of APPLE vs. PEN might change the identity of GREEN [Travis 2001]).

Property 3: *Predicate-argument structure*. One distinctively LoT-like mode of combination is *predication*, in which a predicate is applied to an argument to yield a truth-evaluable structure. Simple sentences like "John smokes" and "Mary is tall" are paradigmatic examples. Other representational formats, such as images and maps, are assessable for accuracy, but often (perhaps always) fail to exhibit truth-evaluable predicate-argument structure (Rescorla 2009; Kulvicki 2015; Camp 2018). We'll usually interpret predicate-argument structure as requiring both discrete constituents and role-filler independence, i.e., as requiring constituents that function as predicates and arguments but maintain type-identity, and as having predicative syntactic structures that can be operated on independently of the content of non-logical constituents. Thus this condition is not merely that the system must be capable of expressing propositions like <John smokes> (a condition that can be met by even the simplest neural nets, where <John smokes> can be represented by an unstructured node), but rather that this predicate-argument structure is instantiated in the representational vehicle itself (see, e.g., Fodor 1987).

Property 4: *Logical operators*. One hallmark of LoT architectures is the use of logical symbols like NOT, AND, OR, and IF. These operators are discrete constituents that compose into larger structures, a hallmark of LoT-like symbols more generally. Logical operators don't obviously presuppose subsentential LoT-like structure, since one could imagine appending such operators to otherwise unstructured formats, or to maps (Rescorla 2009). But they are one piece of an overall LoT-friendly picture, positing discrete constituents that allow for formal-syntactic

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

operations. For example, consider an operation that runs from A-OR-B and NOT-A to B; even if A and B are atomic symbols or maps, their un-LoT-like properties are irrelevant since the operation is sensitive to the logical structure alone. Finding evidence for explicit, discrete logical operators should therefore increase our credence in LoTH, all else equal. We'll construe logical operators as requiring role-filler independence, in that (e.g.) negation operators are the same no matter what proposition they negate.

Property 5: *Inferential promiscuity*. LoT architectures have been useful in characterizing inferential transitions, especially logical inferences (Fodor & Pylyshyn 1988; Rips 1994; Braine & O'Brien 1998; Quilty-Dunn & Mandelbaum 2018a; cf. Johnson-Laird 2006). LoT-like representations should not only encode information, they should be usable for inference in a way that is automatic and independent of natural language.[3] The automaticity point is important: the theories of logical inference just cited share an appeal to computational processes that transform representations with one logical form into representations with another logical form in accordance with rules that are *built into the architecture* (i.e., merely procedural, not explicitly represented, and thus not amenable to intervention from representational states; Quilty-Dunn & Mandelbaum 2018b). If these theories are even roughly on the right track, then we should find evidence for logical-form-sensitive computation outside conscious, controlled, natural-language-guided contexts.

Property 6: *Abstract conceptual content.* LoTH has historically been opposed to concept empiricism, the view that concepts are sensory-based (Barsalou 1999; Prinz 2002). It is logically compatible with other core LoT properties that some LoTs might be modality-specific (e.g., different LoT symbol types and/or syntactic rules for each modality). But there is no a priori reason to expect that primitive LoT symbols—unlike, e.g., iconic or analog formats—will be limited to a certain range of properties (e.g., sensory properties, the referents of simple concepts for classical empiricists). Thus we should expect (*ceteris paribus*) LoT symbols to represent abstract categories without representing specific details (e.g., a symbol that encodes *bottle* and no particular shape or color). There is therefore a non-demonstrative but bidirectional relationship between LoTs and abstract contents: many LoTs should be expected to encode abstract content, and abstract content is naturally represented by means of discrete LoT-like symbols.

***

[3] "Usability for inference" here is independent from structural access constraints, e.g. from modularity.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

The hypothesis that these features cluster together generates non-trivial predictions. Once we've isolated a particular representation-type, evidence for any two features (e.g., discrete constituents and abstract conceptual content) may look completely different. Nonetheless, LoTH predicts that these sorts of evidence should tend to co-occur. This co-occurrence would be surprising from a theory-neutral point of view, but not from the perspective of LoTH. We will use just this sort of clustering-based approach to mount an abductive, empirical argument for LoTH. We focus on independently identified systems to observe whether these six properties cluster in them: perception, physical reasoning in infants and animals, and System-1 cognition.

3. LoT in Computational Cognitive Science

Before we turn to the bulk of our evidence, we first consider the status of LoTH in computational modeling—a topic of pressing concern as the advance of Artificial Intelligence has made LoT appear antiquated to some researchers. LoT-style models naturally grew out of symbolic computation (Fodor 1975; Schneider 2011; cf. Harman 1973; Field 1978), including "GOFAI" ("Good Old-Fashioned Artificial Intelligence": Haugeland 1985). As new computational methods arose that did not presuppose symbolic computation, such as connectionism with its subsymbolic elements, LoT-style architectures grew detractors. With recent successes of subsymbolic deep neural networks (DNNs) (e.g., Google AI's Google Translate, Deep Mind's success with AlphaFold at modeling protein structure and with AlphaZero and MuZero at dominating complex games [Schrittwieser et al. 2020]), LoT-like architectures may appear obsolete.

However, LoT has seen a resurgence in a computational framework that has led to breakthroughs within cognitive science: Bayesianism. Since Bayesian models of cognition are based on probabilistic updating, they appear to present alternatives to LoTH, which posits logical inference. However, Bayesian computational psychology naturally complements LoT architectures (Goodman, Tenenbaum, Feldman, & Griffiths 2008; Kemp 2012; Piantadosi, Tenenbaum, & Goodman 2012; Ullman, Goodman, & Tenenbaum 2012; Erdogan, Yildirim, & Jacobs 2015; Goodman, Tenenbaum, and Gerstenberg 2015; Goodman & Lassiter 2015; Yildirim & Jacobs 2015; Piantadosi, Tenenbaum, & Goodman 2016; Piantadosi & Jacobs 2016; Overlan, Jacobs, & Piantadosi 2017). Wedding probabilistic reasoning to symbolic system processing has led to the "probabilistic language of thought" (PLoT) (Goodman, Tenenbaum, & Gerstenberg. 2015).

PLoTs share a core set of properties: a set of primitives with basic operations for their combination (such as the lambda calculus, e.g., Church from Goodman et al. 2008). Primitives

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

correspond to atomic concepts, which are recursively combined to form concepts of arbitrary complexity (Fodor 1998; Quilty-Dunn 2021). All one must do is define a set of primitives, and a set of rules for combination and the system is capable of constructing a potentially infinite string of well-formed formulae (Chomsky 1965).

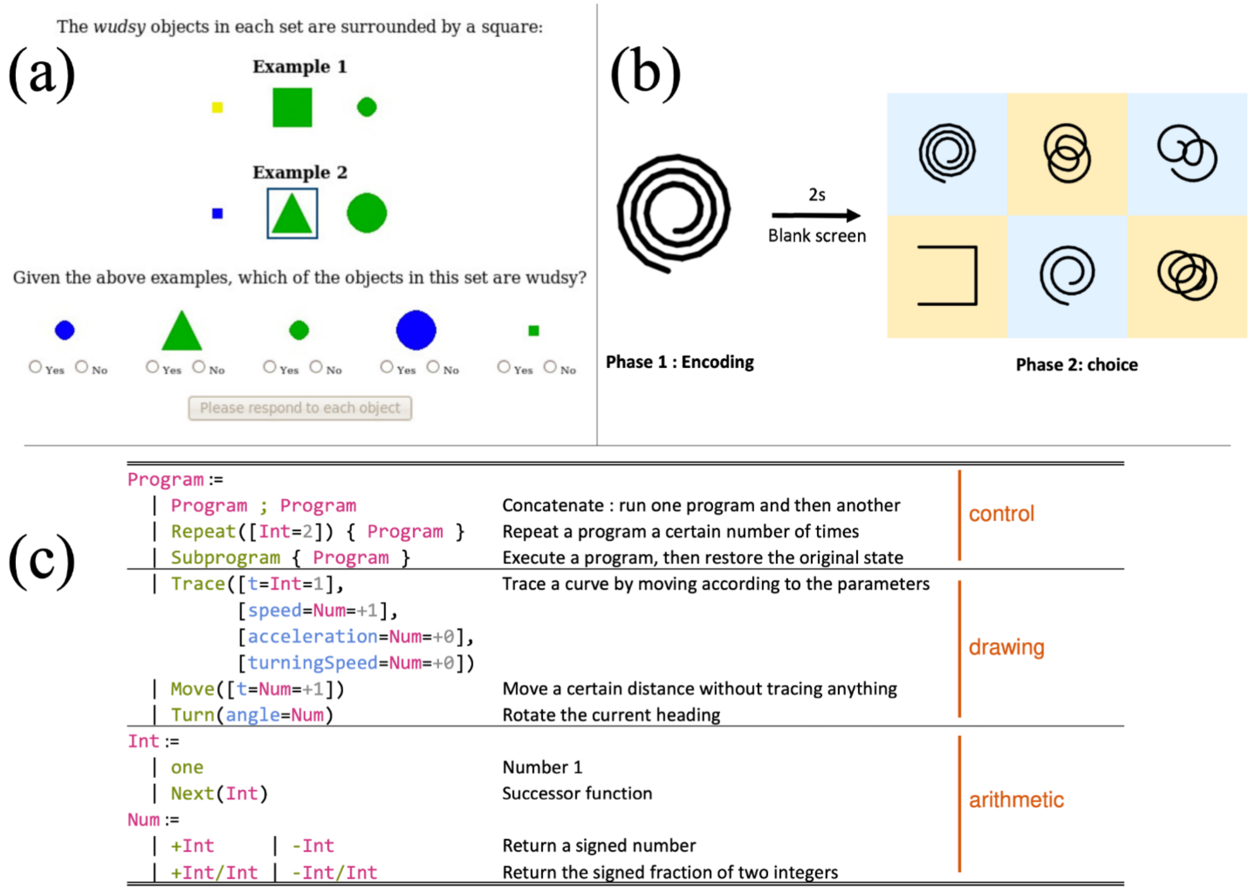
Bayesianism adds probabilistic inference to the traditional LoT machinery. One way of accomplishing this is by having a likelihood function that is noisy (combining this with a preference for simplicity, either because it's explicitly specified as a prior for the system, or because it falls out as a function of other constraints). PLoTs are classical symbolic systems that display all the hallmarks of LoT architectures, such as discrete constituents, role-filler independence, predicate-argument structure, productive and systematic compositionality, and inferential promiscuity. They are also, however, flexible probabilistic computational programs, because all other aspects of symbol processing (e.g., how they are combined, which processes utilize them, which information gets updated for them, even their basic semantics) can be determined probabilistically.

Versions of the PLoT have made serious progress in a number of specific areas, e.g., learning taxonomical hierarchical structures such as kinship (Kemp 2012; Katz, Goodman, Kersting, Kemp, & Tenenbaum 2008; Mollica & Piantadosi 2015), causality (Goodman, Ullman, & Tenenbaum 2011), number (Piantadosi, Tenenbaum, & Goodman 2012), analogical reasoning (Cheyette and Piantadosi 2017), theory acquisition (Ullman, Goodman, & Tenenbaum 2012), programs (Liang, Jordan, & Klein 2010), mapping sentences to logical form (Zettlemoyer & Collins 2005), general Boolean concept learning (Goodman et al. 2008), and moral rule learning (Nichols 2021). The sheer breadth and depth of the Bayesian computational revolution itself provides strong evidence in favor of the viability of the LoT. Instead of computational psychology showing that the LoT is a stale theory of the past, it shows how robust, flexible, powerful, and necessary the LoT is in order to ground our computational cognitive science in a way that maps onto human data.

The models that best approximate one type of human concept learning (e.g., learning that a *wudsy* is the tallest object that is either blue or green) are ones where a fuller set of classical logical connectives are hard-coded as primitives. For instance, Piantadosi et al. (2016) taught participants Boolean and quantificational concepts, then built different LoT models in a lambda calculus and compared them to the human data (Fig. 1a). They found that the models that least resembled human performance tended to have the least LoT-like structure. Models that lacked built-in connectives and represented only primitive features or similarity to exemplars performed poorly, as did models that merely learned response biases and only represented TRUE and FALSE

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

categorization judgments. LoTs built with a single connective from which all others are constructed (such as NAND or conjunctions of Horn clauses, disjunctions with at most one non-negated disjunct) fared better, but not as well as LoTs with the full suite of Boolean operators (conjunction, disjunction, negation, conditional, and biconditional), which in turn were outperformed by models supplanted further with built-in (first-order) quantifiers.[4] While *wudsy* is not an ordinary lexical concept it is a learnable concept for humans and its acquisition is best modeled by a LoT-like architecture. Thus Piantadosi et al.'s findings provide an existence proof for the utility of LoT-like architectures in the acquisition of logically complex, non-lexical concepts.



The *wudsy* objects in each set are surrounded by a square:

Example 1

Example 2

Given the above examples, which of the objects in this set are wudsy?

○ Yes ○ No ○ Yes ○ No ○ Yes ○ No ○ Yes ○ No ○ Yes ○ No

Please respond to each object

```
Program :=
  | Program ; Program                  Concatenate : run one program and then another
  | Repeat([Int=2]) { Program }        Repeat a program a certain number of times
  | Subprogram { Program }             Execute a program, then restore the original state
  | Trace([t=Int=1],                   Trace a curve by moving according to the parameters
          [speed=Num=+1],
          [acceleration=Num=+0],
          [turningSpeed=Num=+0])
  | Move([t=Num=+1])                   Move a certain distance without tracing anything
  | Turn(angle=Num)                    Rotate the current heading
Int :=
  | one                                Number 1
  | Next(Int)                          Successor function
Num :=
  | +Int      | -Int                   Return a signed number
  | +Int/Int  | -Int/Int               Return the signed fraction of two integers
```

Figure 1—(a) Participants draw inferences about the referent of novel terms like *wudsy* based on examples; reprinted from Piantadosi et al. (2016), Figure 1, with permission from American Psychological Association. (b) Participants encode shapes and re-identify them using minimal description length in a PLoT; reprinted from Sablé-Meyer, Ellis, Tenenbaum, & Dehaene

[4] Adding second-order quantifiers did not increase performance, suggesting increasing expressive power *per se* does not necessarily improve model fit.

(2021a), with permission from Mathias Sablé-Meyer. (c) Primitive operations in a geometrical PLoT; reprinted from Sablé-Meyer et al. (2021a), with permission from Mathias Sablé-Meyer.

Bayesian computational psychology provides evidence that we can learn complex concepts by running probabilistic inductions over a distinctive sort of representational system. This system exploits a rich array of discrete constituents (including predicates and logical operators) that compose into predicate-argument structures of the form *A wudsy is an F*; these structures function as inferentially promiscuous hypotheses and incorporate built-in logical operators that obey role-filler independence: in other words, this system is a LoT.[5]

Similar architectures have recently been used to capture representations of geometrical structure (Amalric et al. 2017; Romano et al. 2018; Roumi, Marti, Wang, Amalric, & Dehaene 2021; Sablé-Meyer et al. 2021a; 2021b). For example, Amalric et al. (2017) gave participants a task: observe a sequence of dots and guess where the next dot will appear. They developed a "language of geometry" (see also Romano et al. 2018) and found that the complexity of descriptions in this language predicted human error patterns. Sablé-Meyer et al. (2021a) modified this language (including, e.g., accommodating curve-tracing). Participants took as long as needed to encode shapes, and then re-identified them after a brief delay (Fig. 1b). Description complexity in Sablé-Meyer et al.'s PLoT (Fig. 1c) predicted the duration of both encoding and reidentification.

Our primary aim in this section is to point out that not all cutting-edge computational cognitive science is opposed to LoTH.[6] Indeed, some of the most impressive work in this area relies on LoTs to model human cognition. Current DNNs may be less well-equipped to capture these capacities. For example, Sablé-Meyer et al. (2021b) examined performance of French adults, Himba adults (who lacked formal education or lexical items for geometric shapes and didn't grow up in a "carpentered world"), and French kindergartners on an "intruder" task where they

[5] Bayesian modeling is sometimes pitched as a Marrian "computational-level" rational analysis (Anderson 1990; Oaksford & Chater 2009). However, a model that better captures human behavior than competitors provides defeasible evidence that some approximation of the computational elements of the model are realized in human cognitive architecture. This "algorithmic-level" approach to computational modeling fits with recent Bayesian approaches (e.g., Vul et al. 2014; Lieder & Griffiths 2020). We grant that further evidence is needed to establish the algorithmic-level reality of PLoTs (e.g., behavioral evidence of the sort canvassed in the rest of this paper), but we take their success primarily to push back against the dominance of non-LoT-like architectures such as DNNs. Moreover, the fine-grained behavioral measures used in the "language of geometry" literature discussed in the next two paragraphs evince an algorithmic-level interpretation.

[6] For more critical discussion of DNNs see Lake, Ullman, Tenenbaum, & Gershman 2017 and Marcus 2018.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

had to detect an unusual shape in a crowd of shapes. They found that performance in humans was most similar to a model where shapes are "mentally encoded as a symbolic list of discrete geometric properties" (Sablé-Meyer et al. 2021b, 5). This LoT-like model was contrasted with state-of-the-art DCNNs as well as non-convolutional DNNs (specifically, variational autoencoders), and the LoT model outperformed the alternatives. Furthermore, PLoTs are capable of encoding domain-general models that underwrite commonsensical reasoning, a well-known limitation of extant DNNs (Zhu et al. 2020; Peters & Kriegeskorte 2021). Given the expressive flexibility of PLoTs and their ability to model concept acquisition from just a single data point, they exhibit some advantages over DNN architectures (Piantadosi et al. 2016, 414; cf. Brown et al. 2020; but see Ye & Durrett 2022).

To be clear on the dialectic, many theorists are inclined to point to advances in AI as sufficient evidence against the LoTH. PLoTs serve as an existence proof that LoT architectures are useful in computational modeling. Our claim is not that DNNs will never be able to model this data; indeed, since DNNs are universal function approximators, perhaps such a claim is *ipso facto* false. Other learning policies (e.g., meta-learning; Finn, Yu, Zhang, Abbeel, & Levine 2017) or architectures (e.g., transformers; Vaswani et al. 2017) may turn out to match symbolic models at mimicking acquisition of logically complex concepts and geometrical encoding in humans. We also grant that DNNs are useful for various engineering purposes outside the context of modeling biological competences. Our claim is simply that computational modeling has not left LoT-like symbolic models behind; LoTH remains fruitful in 21st-century computational cognitive science.

It is well-understood by contributors to this literature that "the form that [LoT] takes has been modeled in many different ways depending on the problem domain" (Romano et al. 2018, 2). The PLoTs used to model geometrical cognition possess discrete constituents that combine recursively to form more complex shapes, exhibiting role-filler independence, and encode abstract geometric "primitives" (Amalric et al. 2017) like symmetry and rotation independently of low-level properties. Other PLoTs used to model (complex) concept acquisition possess all these features plus logical operators and predication. Of course, whether any or all of these PLoTs turn out to be isomorphic to human cognition is still—like most questions in cognitive science—open. The two morals we stress are a) that many of these models are meant to test concrete representational formats at the algorithmic level, b) that these models implement LoTs, and (c) that they sometimes match human performance better than competitor models.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

4. Perception

LoTH is often framed as a thesis about thought—that is, post-perceptual central cognition. The idea that perception itself might be couched in a LoT is often ignored (cf. Fodor 1975, Ch. 1; Pylyshyn 2003). Indeed, characterizations of many anti-LoTH views, e.g., concept empiricism, appeal to the hypothesis that conceptual representations have the same format as perceptual representations, implicitly ruling out the possibility of LoT in perception (Prinz 2002; Machery 2016).

We propose instead to take it as an empirical question whether LoT-like representations are deployed in perception, and we'll argue that the answer is likely "Yes". If cognition is largely LoT-like, and perception feeds information to cognition, then we should expect at least some elements of perception to be LoT-like, since the two systems need to interface (Mandelbaum 2018; Quilty-Dunn 2020a; Cavanagh 2021). Our case studies include perceptual representations of objects (e.g., object files), relations within objects (e.g., part-whole relations), and relations between objects.

4.1 Object Files

Object files are perceptual representations that select individuals, track them across time and space, and store information about them in visual working memory (VWM). This construct is probed via independent, but converging methods, including: multiple-object tracking (Fig. 2a; Pylyshyn & Storm 1988), object-based VWM storage (Fig 2b; Hollingworth & Rasmussen 2010), physical reasoning, especially in infants (Fig. 2c; Xu & Carey 1996), and object-specific preview benefits (Fig. 2d; Kahneman, Treisman, & Gibbs 1992). These methods cluster around a common underlying representation, standardly taken to be a unified representational kind (Scholl & Leslie 1999; Carey 2009; Green & Quilty-Dunn 2017; Smortchkova & Murez 2020). Object files are extremely well-studied, are generated by encapsulated perceptual processes (Mitroff, Scholl, & Wynn 2005; Scholl 2007) that operate prior to and independently of natural-language-guided cognition (Carey 2009), and are widely believed to have some sort of compositional structure (minimally, object-property bindings), making them an excellent test-case for LoTH.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

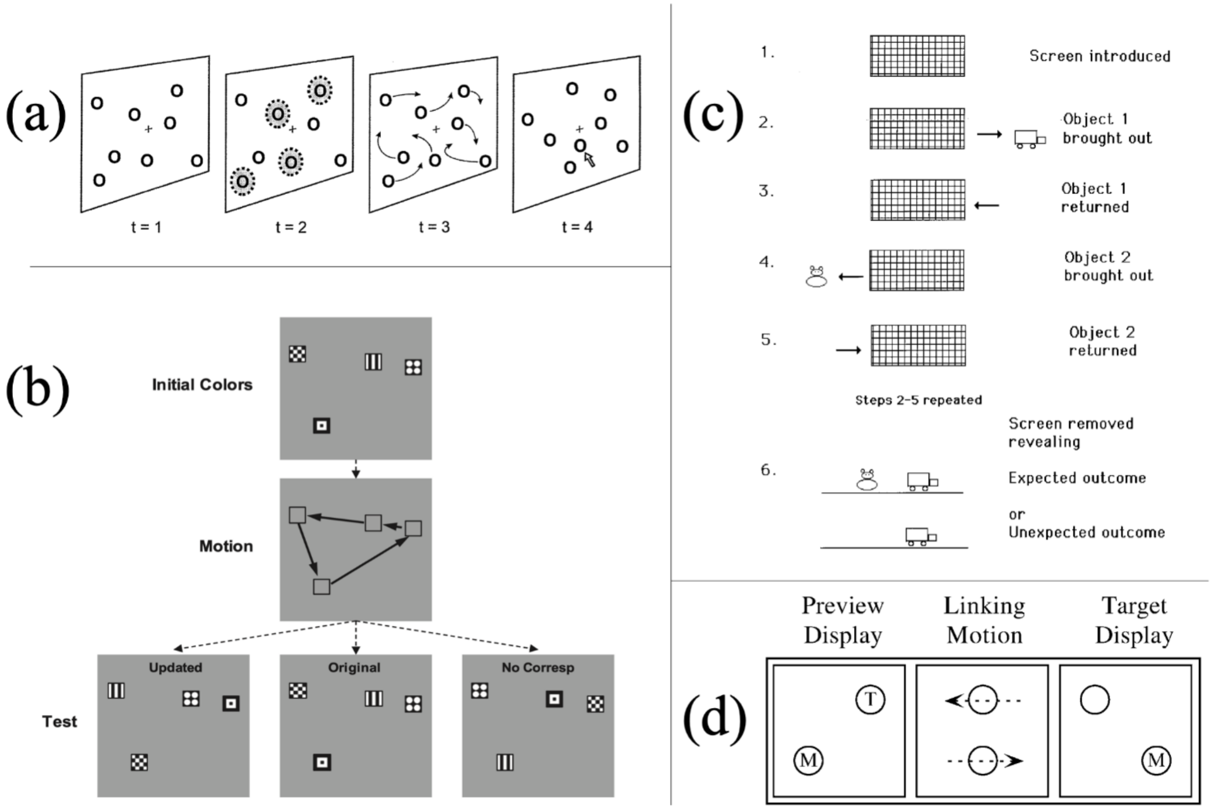


Figure 2. (a) Multiple-object tracking: a subset of visible items (“targets”) are tracked while others (“distractors”) are ignored; reprinted from Pylyshyn (2004), Figure 1, with permission from Taylor & Francis. (b) Object-based VWM storage: a change detection task demonstrates that color is recalled for each object despite location changes, providing just one example piece of evidence that object-based storage in VWM uses object-file representations; reprinted from Hollingworth & Rasmussen (2010), Figure 2, with permission from American Psychological Association. (c) Object-based physical reasoning: objects pop out from behind an occluder, and preverbal infants rely on spatiotemporal information (and featural and categorical information—see Section 5) to keep track of the number of objects, as evidenced by their increased looking time when an unexpected number of items is displayed; reprinted from Xu & Carey (1996), Figure 1, with permission from Elsevier. (d) Object-specific preview benefit: a feature is previewed in each of two visible objects before disappearing, after which the objects move to new locations, and a target feature appears. Subjects show a benefit in reaction time when discriminating the feature if reappears in the same object, illustrating that object-file representations store object properties across spatiotemporal changes; reprinted from Mitroff et al. (2005), Figure 4, with permission from Elsevier.

According to Carey’s (2009) seminal theory of core cognition, object files are amodal but iconic in format (cp. Xu 2019). Nonetheless, we believe a LoT-based model is better suited to the data

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

than an iconic model (Green & Quilty-Dunn 2017; Quilty-Dunn 2020a; 2020c). As far as we know, the possibility of logical operators in object files hasn't been studied. However, converging evidence suggests that object files have discrete constituents, role-filler independence, predicate-argument structure, and abstract conceptual content. In Section 5, we'll explore the inferential promiscuity of object files in physical reasoning.

4.1.1 First, object files exhibit a decomposition into discrete constituents. Unlike rival models (e.g., iconic models), a LoT-based model of object perception predicts that featural representations should easily break apart from (i) representations of individuals and (ii) other featural representations.

Representations of color and shape frequently come apart from representations of objects without disrupting multiple-object tracking (Fig. 2a) (Bahrami 2003; Zhou, Luo, Zhou, Zhuo, & Chen 2010; cp. Pylyshyn 2007). In VWM, object files dynamically lose featural information like color and orientation independently of one another (Bays, Wu, & Husain 2011; Fougnie & Alvarez 2011) and VWM resources are depleted independently for color and orientation (Wang, Cao, Theeuwes, Olivers, & Wang 2017; Markov, Tiurina, & Utochkin 2019). Similar results hold for real-world stimuli. The state of a book (open or closed) is remembered or forgotten independently of its color or token identity (Brady, Konkle, Alvarez, & Oliva 2013), and the identity and state of multiple real-world objects are independently swapped in VWM (Markov, Utochkin, & Brady 2021). These effects are independent of natural-language encoding: they persist when subjects engage in articulatory suppression (Fougnie & Alvarez 2011; Tikhonenko, Brady, & Utochkin 2021), and preverbal infants can lose featural information in VWM but maintain a "featureless" pointer-like component of an object-file (Kibbe & Leslie 2011).

In summary, object files in online tracking and VWM appear to break apart freely into discrete constituents, including representations of individuals and separable feature dimensions. This LoT-like format is independent of natural-language capacities.

4.1.2 Second, object files satisfy demanding constraints on predicate-argument structure. One can grant that object files decompose into discrete constituents but deny that these constituents are ordered into a genuinely sentence-like representation. Here we highlight two constraints on genuinely sentence-like predicate-argument representations: role-filler independence (one of our six LoT properties) and a grammatical attribution/predication distinction.

Recall that role-filler independence requires that discrete constituents compose into larger structures, but the syntactic structure is typed independently of its particular constituents, and the

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

constituents are typed independently of how they happen to compose on a particular occasion. In a predicate-argument structure in particular, both predicate and argument must maintain type-identity independently of their current bindings—e.g., it must be the same JOHN and TALL in TALL(JOHN), TALL(MARY), and SHORT(JOHN).

The clear candidates for predicate-like and argument-like representations in object files are representations of properties and representations of individuals, respectively (cp. Cavanagh 2021). Representations of individuals must maintain their identity independently of the properties they bind, since tracking performance is successful while properties change (Flombaum, Kundey, Santons, & Scholl 2004; Flombaum & Scholl 2006; Zhou et al. 2010) and even while properties are forgotten entirely (Scholl, Pylyshyn, & Franconeri unpublished; Bahrami 2003). The computational processes involved in tracking are known as object correspondence processes. Some properties are used to compute object correspondence (e.g., spatiotemporal features and some surface features—see below). However, the fact that the argument-like representation of the tracked individual can persist while many attributed features are changed/lost entails that the representation maintains independence from the properties to which it is bound.

Likewise, representations of properties maintain their identity independently of the object-representations to which they're bound. Some evidence for this is the already-cited fact that they regularly come apart from their respective object representations. However, more striking evidence comes from the way in which featural information is "swapped" between objects. Participants often misremember a feature of one object as bound to another object (Bays, Catalao, & Husain 2009), including for real-world stimuli (Utochkin & Brady 2020; Markov et al. 2021). Even during multiple-object tracking, a stored feature of one object (e.g., a previewed numeral) may be swapped with another object if they come too close to each other during tracking (Pylyshyn 2004). Thus property-representations, like individual-representations, maintain type-identity across distinct bindings, demonstrating role-filler independence.

The second constraint on predicate-argument structure is a grammatical attribution/predication distinction. In a genuinely sentence-like representation, we can distinguish grammatical positions of predicates. For example:

(1) That spherical object is red.
(2) That red object is spherical.

Both attribute spherical shape to the referent of "That", but in (1) the predicate falls within the scope of the noun phrase, while in (2) it is in main-predicate position.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

One way of capturing this distinction is by appeal to the role of the predicate in grounding the reference of the noun phrase. For example, Perner and Leahy characterize thought in terms of file-like representations (cp. Recanati 2012), which “capture the predicative structure of language, i.e., the distinction between what one is talking about (the subject, topic, i.e., what the file tracks) and what one says about it (the information about the topic, i.e., the information the file has on it)” (2016, 494). Files have “labels” that are captured by (inter alia) determiner phrases like THE RABBIT as well as file-contents that include predicates like +FURRY. The attribution of RABBIT in THE RABBIT plays some reference-grounding role, while +FURRY is parasitic on the referent of THE RABBIT and merely predicates a property of that referent (see Burge 2010). In particular, the label-like attributive helps to sustain, and constrain, reference of the file over time.

We can exploit the attribution/predication distinction to see whether the discrete constituents of object files are organized in a genuinely predication-like way, or whether they are merely label-like representations, as in THE RABBIT. The latter format is compatible with a LoT-based model, but part of the virtue of LoTH is that it predicts nontrivial clustering of LoT-like properties. We ought to predict full-blown propositional structures are present in perception as well.

Object files attribute a wide range of properties to their referents, and some of these are used to guide reference to objects. For example, an object file will continue to refer to an object that disappears behind an occluder, but only if it re-emerges at a spatiotemporally appropriate location (Scholl & Pylyshyn 1999). However, while object files attribute other features like color, reference to the object is maintained even if it re-emerges a totally different color. Generalizations like this have led some researchers to describe spatiotemporal features as aspects of the object-file “label” while surface features are “stored inside the folder” (Flombaum, Scholl, & Santos 2009, 153). Recent evidence casts doubt on strict limitations on which properties are part of the “label”. While earlier theories took spatiotemporal indices to be uniquely privileged (e.g., Leslie et al. 1998), surface features like color can play an indexing, reference-guiding role in object files, even in ordinary contexts (Hollingworth & Franconeri 2009; Moore, Stephens, & Hein 2010; Hein, Stepper, & Moore 2021). However, object files routinely store some featural information (e.g., color or orientation) while completely failing to use it to guide reference to objects (e.g., Gordon, Vollmer, & Frankl 2008; Richard, Luck, & Hollingworth et al. 2008; Gordon & Vollmer 2010; Jiang 2020; see Quilty-Dunn & Green forthcoming for a review).

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Object files not only contain discrete constituents, but also the way those constituents are organized satisfies demanding criteria for predicate-argument structure.

4.1.3 Third, object files encode abstract conceptual content. Part of the utility of LoT-like formats is abstracting away from modality-specific information. A LoT allows color and categorical information to be captured in the same representation, as in THAT OBJECT IS A BROWN RABBIT. If object files are LoT-like representations, they not only ought to encode conceptual categories, they ought to do so in a way that abstracts away from sensory details.

The evidence suggests that object files do encode abstract conceptual content. For example, the object-specific preview benefit—a reaction-time benefit in discriminating previously viewed properties of tracked objects (Fig. 2d)—is observed even when the previewed feature is an image of a basic-level category (e.g., APPLE) and the test feature is the corresponding word (e.g., "apple") (Gordon & Irwin 2000). Similar effects are found for semantic identity of words across fonts (Gordon & Irwin 1996) or basic-level categories across different exemplars (Pollatsek, Rayner, & Collins 1984) and across visual and auditory information (Jordan, Clark, & Mitroff 2010; cf. O'Callaghan forthcoming). Importantly, these effects do not transfer across associatively related stimuli (e.g., bread-butter), ruling out a reductive associative explanation (Gordon & Irwin 1996).

Similar effects were recently found in preverbal infants. Kibbe and Leslie (2019) discovered that while infants will not notice whether the first of two serially hidden objects changes its surface features when it re-emerges from behind an occluder, they do notice when it changes its category between FACE and BALL. Pomiechowska and Gliga (2021) tested preverbal infants in an EEG change-detection task for familiar categories (e.g., BOTTLE) or unfamiliar categories (e.g., STAPLER). Infants showed an equal response in the negative-central ERP (an EEG signature of sustained attention) for across-category and within-category changes for unfamiliar categories, suggesting, unsurprisingly, failure to categorize. But for familiar categories, they showed an increased amplitude only for across-category changes, suggesting that their object files in VWM maintained the conceptual category of the object while visual features decayed.

In adults, VWM seems often to discard specific sensory information in favor of conceptual-category-guided representations (Xu 2017; 2020; cf. Harrison & Tong 2009; Gayet, Paffen, & Van der Stigchel 2018). Participants recall blurry images as less blurry than they really were, suggesting categorical encoding that "goes beyond simply 're-experiencing' images from the past" (Rivera-Aparicio, Yu, & Firestone 2021, 935). Bae, Olkkonen, Allred, & Flombaum (2015) found that object files in online perception and VWM are biased toward the center of color

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

categories, suggesting that object files store a basic-level color category like RED plus a noisy point estimate within the range of possible red shades. This evidence implicates a category-driven format for object-based VWM representations that abstracts away from low-level visual detail.

Object files encode abstract conceptual content in a way that is not reducible to low-level modality-specific information, just as a LoT-based model predicts.

4.2. Structured relations

We've just argued that perceptual representations of individual objects contain discrete constituents that are organized in a predicate-argument structure and predicate abstract conceptual contents—in other words, they're sentences in the LoT. We'll now describe some LoT-like properties of representations used in the perception of structured relations, both within and between objects.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

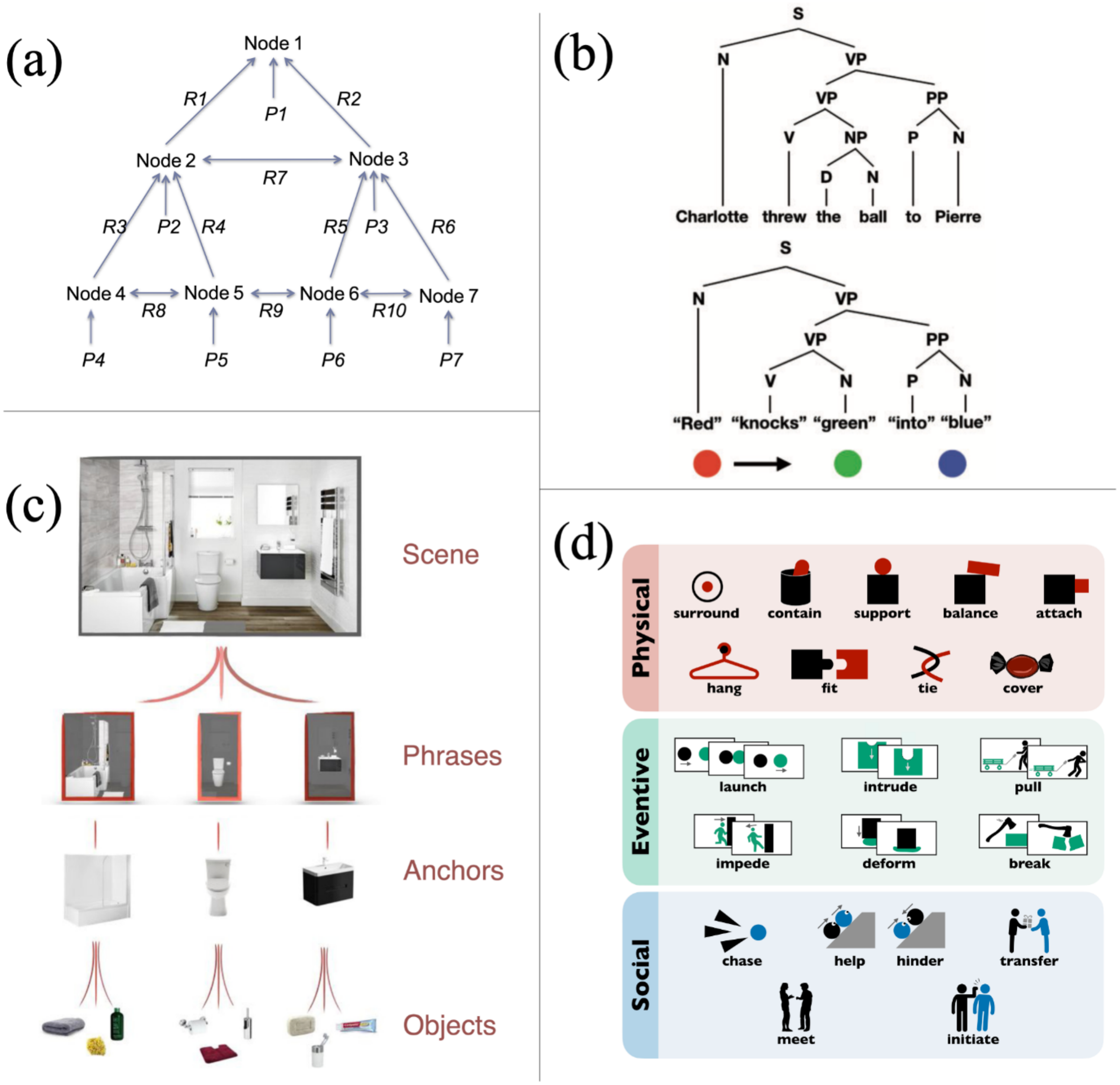


Figure 3. (a) Hierarchical part-whole structural description: Ps=monadic featural properties, horizontal Rs=spatial relations, vertical Rs=mereological relations; reprinted from Green (2019), Figure 9, with permission from Wiley. (b) Structural analogy between tree-like structures in natural language syntax and tree-like perceptual representations of interobject relations; reprinted from Cavanagh (2021), Figure 3, SAGE Publishing under CC BY 4.0, cropped and rearranged. (c) Hierarchical structure in scene grammar: objects are organized relative to "anchors" (relatively large, immobile elements of environments like showers and trees) in phrase-like structural descriptions of normal relative positions; reprinted from Võ et al. (2019), Figure 2, with permission from Elsevier. (d) Examples of perceived interobject relations; reprinted from Hafri & Firestone (2021), Figure 2, with permission from Elsevier.

4.2.1 First, our perceptual systems represent hierarchical part-whole structure. Our perceptual systems don't simply select objects and attribute properties to them. They also break objects down into component parts and represent their part-whole structure. When we perceive a pine tree, we see a branch as part of the tree and a needle as part of the branch, with a sense of the borders between these various parts. Thus the visual system makes use of hierarchical structural descriptions (Fig. 3a; Hummel 2013; Green 2019).

The motivation for classic structural-description accounts of object perception was computational: positing representations of object parts that compose to generate descriptions of part-whole structure allows for successful computational modeling of object perception (Marr & Nishihara 1978; Biederman 1987). These models operate just as a classical LoT picture demands, exhibiting systematic and productive compositionality of viewpoint-invariant descriptions of parts (Fig. 3b; Cavanagh 2021). Structural descriptions "are compositional—forming complex structures by combining simple elements—and thus meaningfully symbolic" (Saiki & Hummel 1998b, 1146).[7]

One of the key assumptions of such models is that object-part boundaries are psychologically real, i.e., two points will be treated differently by the visual system when they lie on the same part as opposed to two different parts of the same object. This assumption turns out to be true (Green 2019). For example, a well-known example of object-based attention is that two stimuli are better discriminated when they lie on the same object than different objects, controlling for distance (Duncan 1984; Egly, Driver, & Rafal 1994). The same is true within parts of objects: participants are quicker to discriminate targets if they lie on the same part than if they cross a part-boundary (Barenholtz & Feldman 2003). Furthermore, unfamiliar object pairs that share structural descriptions are seen as more similar than object pairs that have a higher degree of overall geometrical similarity but different structural descriptions (Barenholtz & Tarr 2008).

Role-filler independence emerges directly from structural description models, often explicitly so (Hummel 2000). Some independent evidence comes from Saiki and Hummel (1998a), who found that shapes of parts and their spatial relations are not represented holistically—in other words, the type-identity of each part is represented independently of its particular role in the structural description and vice versa. Similarity judgments are also guided independently by part-shapes and their interrelations, suggesting role-filler independence (Goldstone, Medin, & Gentner 1991).

[7] It's possible that "skeletal" shape representations (Feldman & Singh 2006; Firestone & Scholl 2014) exhibit similar LoT-like structure (Green ms).

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

We don't deny that the visual system also employs holistic view-based template-like representations (Ullman 1996; Edelman 1999) and other formats. Our claims are merely (i) structural descriptions are among the many representations used in visual processing, and (ii) they have a LoT-like format comprising discrete constituents ordered in hierarchical ways that preserve role-filler independence (Fig. 3b).

4.2.2 Second, we perceive structured relations between objects. We don't perceive objects as isolated atoms, as if through a telescope. Instead, we see the glass on the table, the pencils in the cup, etc.

In a recent review, Hafri and Firestone (2021) survey striking evidence that such relations are recovered rapidly and in abstract form in visual processing (Fig. 3d). For example, the visual system distinguishes containment-events (one object disappears inside another) from occlusion-events (one disappears behind another) (Strickland & Scholl 2015). A hallmark of categorical perception is greater discrimination across than within category-boundaries; participants are better at identifying changes in the position of two circles if the change places the circles in a distinct relation (e.g., CONTAIN(X,Y), TOUCH(X,Y), etc.), suggesting categorically perceived interobject relations (Lovett & Franconeri 2017). When participants are searching for a particular relation like cup-contains-phone, they are more likely to have a "false-alarm" for target images that instantiate the same relation, like pan-contains-egg, but not book-on-table (Hafri, Bonner, Landau, & Firestone 2021).

Like structural descriptions, perceptual representations of abstract relations exhibit role-filler independence. Abstract relations apply independently of the relata, and representations of relata persist once the relation is broken—e.g., it's the same ON in ON(CAT,COUNTER) and ON(KETTLE,STOVE), and it's the same CAT once the cat leaps off the counter. Hafri et al.'s (2021) finding is especially relevant: the relation CONTAIN(X,Y) governs similarity judgments independently of the relata, about as clear a demonstration of role-filler independence as one could expect to find.

It would be efficient for the visual system to store frequently represented relations. A fascinating recent literature on "scene grammar" (Fig. 3c; Võ 2021; Kaiser, Quek, Cichy, & Peelen 2019) details effects of representations of structured relations in visual long-term memory on visual search (Draschkow & Võ 2017), categorization (Bar 2004), consciousness (Stein, Kaiser, & Peelen, 2015), and gaze duration (Võ & Henderson 2009). Relational representations in visual long-term memory (e.g., ON(POT,STOVE)=yes, IN(SPATULA,MICROWAVE)=no) aren't based on associations or statistical summaries over low-level properties. They persist despite

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

changes in position and context (Castelhano & Heaven 2011), thus abstracting away from overlearned associations. Characteristic scene-grammar effects disappear, however, for upside-down stimuli (Stein et al. 2015), implicating a categorical rather than low-level format. The effects also appear not to rely on summary-statistical information represented outside focal attention (Võ & Henderson 2009). Despite developing independently of natural language (Öhlschläger & Võ 2020), structured relations in scene grammar display curious hallmarks of language-like formats. For instance, the P600 ERP increases for syntactic violations in language, and also increases for stimuli that violate visual scene "syntax" (e.g., mouse-on-computer instead of mouse-beside-computer; Võ & Wolfe 2013). It's standard to talk of scene grammar as associative, but its relational components satisfy a handful of our LoT hallmarks (e.g., discrete constituents with role-filler independence that encode abstract contents, including categories and relations, and function as arguments in multi-place predicates as in ABOVE(MIRROR,SINK)). Scene grammar is used directly in controlled behavior (e.g., how to arrange a VR scene; Draskchow & Võ 2017); how broadly it can function in logical inference remains to be explored experimentally.

4.3 Vision and DNNs

In sum, our perceptual capacities to identify and track objects, grasp their characteristic structures, and perceive and store their relations with one another, appear to rely on LoT-like representations.

A major source of contemporary skepticism about LoTH is the rise of DNNs. Apart from large language models like GPT-3, nowhere are DNNs more visible as models of human cognitive capacities than in visual perception. Given their successes at image classification and apparent similarities to biological vision, one might wonder whether the subsymbolic network structure of DNNs obviates the need to posit LoT-like structures.

The DNNs that have been most touted as models of biological vision are deep convolutional neural networks (DCNNs) trained to classify images (Kriegeskorte 2015; Yamins & DiCarlo 2016). After training on large data sets like ImageNet, DCNNs exhibit remarkable levels of performance on image classification. It is important to evaluate comparisons to human vision not simply in terms of performance, but primarily in terms of underlying competence (Chomsky 1965). Just as differences in performance need not entail differences in competence (Firestone 2020), human-like performance on a limited range of tasks need not entail human-like underlying competence. In other words, DCNNs may accomplish image classification while lacking key structural features of human vision, including those relevant to LoTH.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

DCNNs have been argued to resemble primate vision in competence as well as performance by appeal to metrics of similarity such as "Representational Similarity Analysis" (Khaligh-Razavi & Kriegeskorte 2014) and "Brain-Score" (Schrimpf et al. 2018). However, there are shortcomings both to earlier findings of high similarity using these metrics and to the metrics themselves. For example, Xu and Vaziri-Pashkam (2021b) used higher quality fMRI data for their Representational Similarity Analysis and found that, contra Khaligh-Razavi and Kriegeskorte's earlier findings, high-performing DCNNs (both feedforward and recurrent) show large-scale dissimilarities to human vision. Brain-Score has been criticized for insufficient sensitivity to architectural distinctions (e.g., feedforward vs. recurrent models): "either the Brain-Score metric or the methodology with which a model is evaluated on it fails to distinguish among what we would think of as fundamentally different types of model architectures" (Lonnqvist, Bornet, Doerig, & Herzog 2021, 3). Furthermore, while Schrimpf et al. (2018) found that Brain-Score positively correlates with image classification performance, it fails to capture the crucially hierarchical structure of human vision. Nonaka, Majima, Aoki, & Kamitani (2021) thus developed a "Brain Hierarchy Score" that measures similarities between hierarchical structures, applied it to 29 DNNs, and found a negative correlation between image classification performance and similarity to human vision. This finding provides a striking illustration of how DNNs can excel in performance while veering apart from human competence (see also Fel, Felipe, Linsley, & Serre 2022).

Our case for LoT in vision is limited to certain domains: objects, relations between parts and wholes, and relations between objects. It is not a coincidence, in our view, that DNNs that succeed at image classification exhibit little to no competence in these domains. As Peters and Kriegeskorte write about feedforward DCNNs, "the representations in these models remain tethered to the input and lack any concept of an object. They represent things as stuff" (2021, 1128).[8] It is also not clear that DCNNs are capable of representing global shape, let alone the relation between global shape and object-parts (Baker & Elder 2022). Baker, Lu, Erlikhman, & Kellman (2020) trained AlexNet, VGG-19, and ResNet-50 to classify circles and squares, but found that these DCNNs relied only on local contour information; circles made of jagged local edges were classified as squares, and squares made of round local curves were classified as circles. The same models (and several others) also could not distinguish possible from impossible shapes, which requires relating local contour information to global shape (Heinke, Wachman, van Zoest, & Leek 2021). Failures at processing relations hold not only for DNNs that map images to labels, but also those that map labels to images: Conwell and Ullman (2022) fed the text-guided

[8] Of course DNNs trained for multiple-object tracking do much better (Xu et al. 2019; Burgess et al. 2019), but their similarity to human visual competence is underexplored.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

image-generation model DALL-E 2 a set of interobject relations (including those used by Hafri et al. [2021]) and found that it failed reliably to distinguish, e.g., “a spoon in a cup” from “a cup on a spoon”.

To be clear, we make no claims about in-principle limitations of DNNs. The machine-learning literature is extremely fast-moving, and we do not pretend to know what it will look like in even one year’s time. Moreover, different DNN architectures might better capture the visual processes discussed here. While convolutional architectures might privilege local image features, perhaps non-convolutional architectures like vision transformers (Vaswani et al. 2017) are better suited to avoid these limitations and will supersede DCNNs as models of human vision (Tuli, Dasgupta, Grant, & Griffiths 2021). Since DCNNs have accumulated enormous publicity despite apparently lacking basic elements of biological vision like global shape and objecthood, future DNN-human comparisons should be approached with caution. Finally, as was noted long ago, neural-network architectures might be able to implement a LoT architecture (Fodor & Pylyshyn 1988). Indeed, some recent work on DNNs explores implementations of variable binding (Webb, Sinha, & Cohen 2021; though see Gröndahl & Asokan 2022; Miller, Naderi, Mullinax, & Phillips 2022), a classic example of LoT-like symbolic computation (Marcus 2001; Gallistel & King 2009; Green & Quilty-Dunn 2017; Quilty-Dunn 2021). Our six core LoT properties help specify a cluster of features that such an implementation should aim for.

DNNs are marvels of contemporary engineering. It does not follow that they recapitulate architectural aspects of human vision. We agree with Bowers et al.’s (2022) recent complaint that research on DNNs as models of biological vision is overly focused on performance benchmarks and insufficiently guided by experimental perceptual psychology. Given that DNNs are universal function approximators, and given the vast resources being poured into their development, they will only get closer to human performance over time. But this performance will not reflect core competences of the human visual system unless the relevant models incorporate LoT-like representations of objects and relations.

5. LoTs in Non-Human Animals and Children

Traditionally, theorists in animal and infant cognition have been reluctant to posit complex cognitive processes, let alone computations over LoT-style representations (e.g., Morgan 1894; Premack 2007; Penn, Holyoak, & Povinelli 2008; cf. Fitch 2019). However, the state of the art in comparative and developmental psychology is surprisingly congenial to LoTH.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

5.1 Abstract Content and Physical Reasoning

Considerable evidence suggests infants use object files to reason about the identity, location, and numerosity of hidden objects (Spelke 1990; Carey 2009). However, in a foundational study, Xu & Carey (1996) found that, while 12-month-olds who see a duck and then a ball pop out from behind an occluder expect two objects to be present, 10-month-olds don't. This failure might seem to suggest that abstract conceptual content is not usable for physical reasoning in young infants, potentially undermining LoT-based models of infant reasoning (Xu 2019).

However, 10-month-olds do succeed for socially significant categories (Bonatti, Frot, Zangl, & Mehler 2002; Surian & Caldi 2010) and objects that are made communicatively salient (Futo, Teglas, Csibra, & Gergely 2010; Xu 2019, 843). There is also evidence that priming can allow infants to use information in physical reasoning many months earlier than they would otherwise appear to. Lin et al. (2021) made features (e.g., color) salient by first showing an array of objects that differed along the relevant dimension (e.g., all different colors). This nonverbal priming allowed infants to use information in object files to reason about the individuation of hidden objects six months earlier than other methods had detected (e.g., while infants had not shown surprise at a lop-sided object balancing on a ledge until 13-months, Lin et al.'s nonverbal priming of lop-sidedness caused seven-month-olds to show the effect).

Infants should therefore be able to use conceptual categories for Xu and Carey's individuation task long before 12-months if the right information is primed first: e.g., the relevance of the category's function, a key aspect of artifact concepts (Kelemen & Carey 2007; cf. Bloom 1996). Stavans and Baillargeon (2018) demonstrated objects' characteristic functions before hiding (Fig. 4a) and found four-month-olds succeeded at Xu and Carey's individuation task, looking longer when only one object was revealed. These results show two key LoT-like features—abstract content and inferential promiscuity—in extremely young preverbal infants. Thus the earlier failures seem to be explained by performance constraints (Stavans, Lin, Wu, & Baillargeon 2019).

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

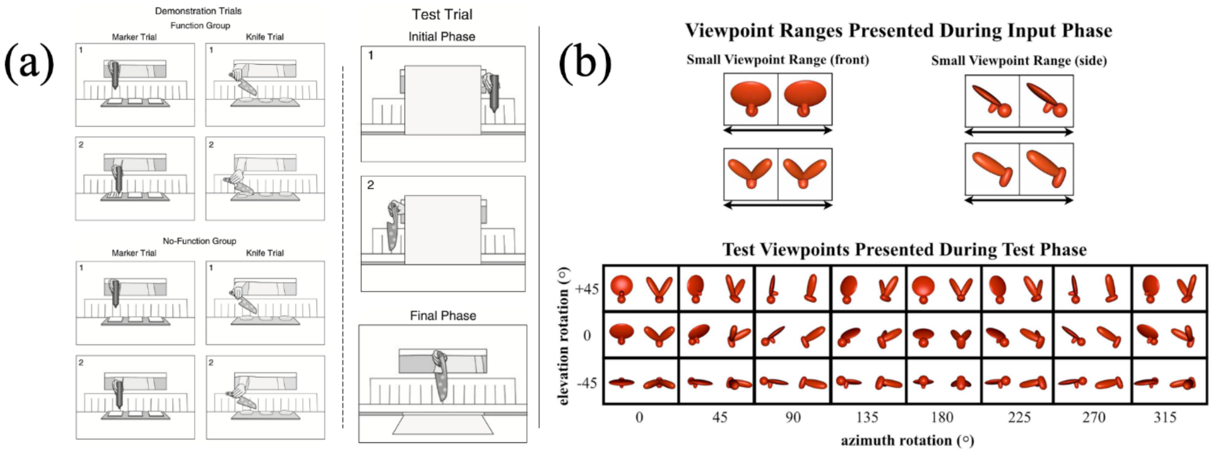


Figure 4. (a) Function demonstrations aid object individuation: in a modification of Xu & Carey's (1996) paradigm, infants first see the characteristic function of an object demonstrated (e.g., a marker drawing, a knife cutting), and this demonstration primes them to use categorical and featural information about the objects to expect two objects in the test trials (i.e., increased looking time when only one object appears); reprinted from Stavans & Baillargeon (2018), Figures 4 and 5, with permission from Wiley. (b) View-invariant information extracted by newborn chicks: chicks are shown a highly limited set of viewpoints on an object and form an abstract, view-invariant representation; reprinted from Wood & Wood (2020), Figure 1, with permission from Elsevier.

The use of abstract content in physical reasoning is arguably present throughout the animal kingdom, and is well-studied in primates (e.g., Flombaum et al. 2004) and even some arthropods. Loukola, Perry, Coscos, & Chittka (2017) trained bumblebees through social learning (using a dummy-bee) to roll a ball—an unusual behavior for bumblebees in the wild—into the center of a platform for a sucrose reward. When the platform was later re-arranged with several balls at various locations that the bees could push into that central area, the bees opted to push balls closest to the center of the platform, even if they differed in color or location from the one they had seen pushed initially. This suggests bumblebees are sensitive to shape in a way that is dissociable from color and location, in contrast to many model-free learning accounts but just as one would expect if shape-type is encoded in a LoT. In a similar vein, Solvi, Al-Khudhairy, & Chittka (2020) found that bumblebees could recognize objects under full light that they had previously encountered only in darkness, suggesting they can transfer shape representations stored through touch to a visual task. Bumblebees therefore appear to represent shape in a way that is dissociable from modality-specific low-level features. These representations figure in practical inferences (thereby displaying inferential promiscuity), and that guides recognition across modalities (thereby displaying abstract content). Furthermore, honeybees trained on a *fewer-than* relation (e.g., 2<5) were able to generalize to cases involving zero items (e.g., 0<6)

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

without any zero-item training, implicating an abstract symbolic representation of *zero* that guides inferential generalization and logico-mathematical reasoning (Howard, Avargues-Weber, Garcia, Greentree, & Dyer 2018; cf. Vasas & Chittka 2019; see Weise, Cely Ortiz, & Tibbets 2022 for abstract contents of same and different). Similarly, bees' navigational inferences have been used as an argument for a bee LoT because of their computational complexity (Gallistel 2011).

Much of our discussion in Sections 4 and 5.1 has concerned abstract (e.g., amodal or view-invariant) object representations, and one might wonder whether these effects are really due to associations between low-level features acquired gradually during development. One might therefore wonder whether DNNs could therefore provide a better explanation of these effects. However, Wood and Wood (2020) found that newborn chicks showed one-shot learning of abstract object representations (fig 4b). Shortly after birth, having been reared in an environment with no movable-object-like stimuli, chicks were shown a virtual 3D-object rotating either fully 360-degrees, or just 11.25-degrees; later, the chicks successfully recognized the objects from arbitrary viewpoints (equally well in both conditions) and moved towards them. Given the paucity of relevant input, this experiment points away from DNN-based explanations of abstract object representations.

Similarly, Ayzenberg and Lourenco (2021) showed preverbal infants a single view of 60-degrees of an unfamiliar object; using a looking-time measure, they found that the infants formed an abstract, categorical representation, recognizing the object even when viewpoint and salient surface features had drastically changed. The infants' one-shot category learning outperformed DCNNs trained on millions of labeled images. This divergence between DCNN and human performance echoes independent evidence that DCNNs fail to encode human-like transformation-invariant object representations (Xu & Vaziri-Pashkam 2021a).

5.2 Logical Inference

Proponents of LoTH have long held up its ability to explain logical inference in pre-verbal children and non-human animals as a virtue (Fodor 1983; Fodor & Pylyshyn 1988; Cheney & Seyfarth 2008; Gallistel 2011; cf. Bermudez 2003; Camp 2007, 2009; Gauker 2011). Recent evidence suggests infants and animals may use logical operators in logical inferences.

Consider the growing body of work on disjunctive syllogistic reasoning (DS). A standard means of testing for this capacity is Call's (2004) two-cups task. The task involves placing a reward in one of two cups behind an occluder. Once the cups are brought back into plain view, the

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

participant is shown that one is empty, and can then choose which of the two cups to select from. Typically, researchers are interested in whether the participant selects the unrevealed cup more often than the revealed one, and whether they choose it without inspecting it first. Such behavior is often taken as evidence the participant can reason through DS, since there's definitely a reward, and one of the two cups is empty, guaranteeing the location of the reward by DS. A surprising number of animals succeed at this task, as well as children as young as two (Call 2006).

Mody & Carey (2016) argue that there is a confound in such tasks. Participants could rely on a non-logical strategy involving modal operators: They could form two unrelated beliefs, MAYBE THERE IS A REWARD IN CUP A and MAYBE THERE IS A REWARD IN CUP B. On this strategy, once shown that cup A is empty, participants simply ignore the possibility that there may be a reward there; left only with the belief that there may be a reward in cup B, they then select cup B. So the authors modified this task, using two rewards and four cups (Fig. 5a). While children as young as 2.5 succeed at the two-cup task, only 3- and 5-year olds succeed at this four-cup task, with 5-year olds performing best.

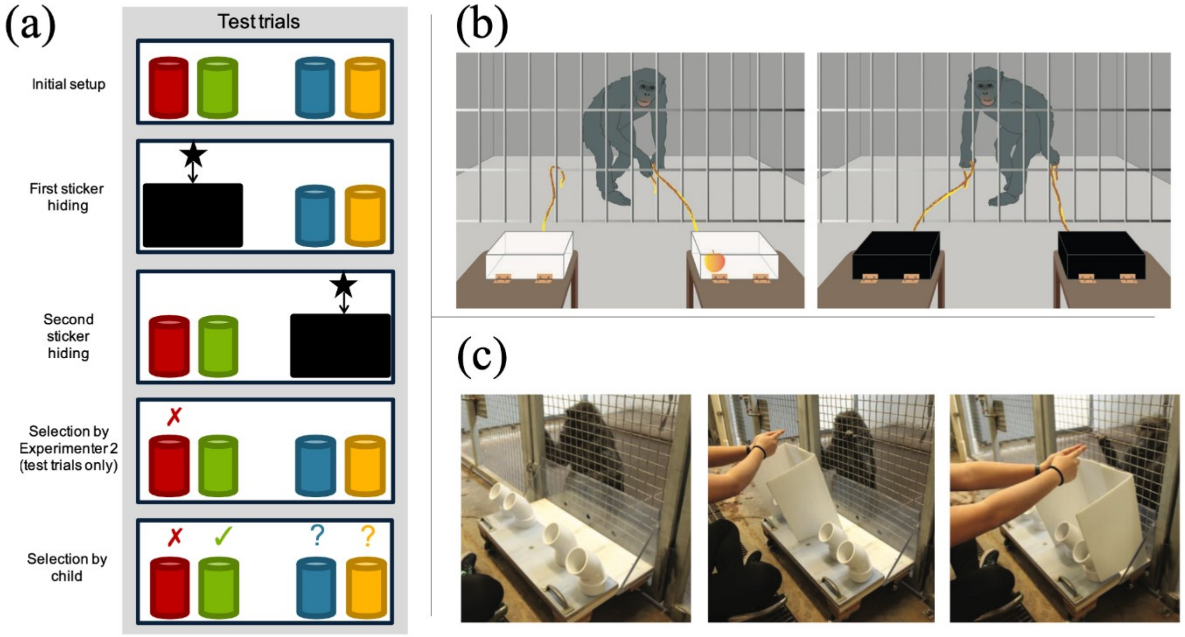


Figure 5. (a) Four-cup task: a reward is placed behind an occluder and into one of two cups, and again for another reward and pair of cups. Then one cup is shown to be empty, and participants who perform disjunctive syllogism can infer that a reward is certain to be in the other cup in that pair; reprinted from Mody & Carey (2016), Figure 1, with permission from Elsevier. (b) Alternatives in chimps: a reward is placed in one of two boxes, and chimps pull a string to open

the box and reveal the reward. The chimps pull both boxes when they are opaque, suggesting simultaneous representation of two possibilities; reprinted from Engelmann et al. (2021), Figure 1, with permission from Elsevier. (c) Success on four-cup task by baboons, reprinted from Ferrigno et al. (2021), Figure 1, SAGE Publishing.

Pepperberg, Gray, Mody, & Carey (2019) found that an African grey parrot, Griffin, succeeded at a modified version of the four-cup task. Remarkably, Griffin selected the cup that contained reward (a cashew) on nearly every trial (chance, in this case, was 33%), besting human five-year-olds (whose success is surprisingly variable; Gautam, Suddendorf, & Redshaw 2021). More moderate success at the four cup task has also been achieved with olive baboons (Fig. 5c; Ferrigno, Huang, & Cantlon 2021).

A straightforward way of understanding these results is to accept that at least some non-human animals are competent with DS. To execute that inference, one needs two sentential connectives, NOT and OR. These must be combined, syntactically, with representations of states of affairs.

The failure of younger kids at Mody and Carey's 4-cup task at first looks like bad news for LoTH. However, it might only reflect a failure with using negation, rather than with logical inference more broadly (Feiman, Mody, & Carey 2022). Moreover, as with Xu's (2019) arguments against LoT-like format in object files, the possibility of performance demands masking an underlying LoT-based competence is plausible. The 4-cup task requires kids to track four cups divided into two pairs and two occluded stickers, which is demanding on VWM; indeed, animals who outperform children tend to have superior VWM capacity (Pepperberg et al. 2019, 417; cf. Cheng & Kibbe 2021). As Pepperberg et al. point out, younger children also act more impulsively than older ones, sometimes ignoring relevant knowledge in demanding tasks. Thus we should look for less demanding tasks before ruling out LoT-like logical inference in children. For example, we could look for independent psychophysical signatures of DS as performed by adults and see whether those signatures are present in children in simpler tasks.

Cesana-Arlotti et al. (2018) showed 12-month-olds and adults two objects hidden behind occluders (e.g., a snake and ball); they saw one placed in a cup without knowing which, and finally the unmoved object (e.g., snake) popped out, allowing subjects to infer the identity of the cup-hidden object (ball). When the cup-hidden object was revealed, infants' looking time showed they expected it to be the yet-unseen object (ball). This finding is compatible with non-logic-based explanations. However, Cesana-Arlotti et al. found that adults performing DS showed an oculomotor signature: during inference, their pupils dilated and eyes darted to the still-hidden

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

object. This same signature was found in the infants, implicating the same underlying computations.

Genuine DS should be domain-general. Cesana-Arlotti, Kovacs, & Téglás (2020) used a similar paradigm to test DS in twelve-month-olds, this time relying on their knowledge of others' preferences. Participants learned an agent's preference among objects (ball vs. car); the non-preferred object then briefly popped out from behind its occluder, after which the agent reached behind one of the occluders. Twelve-month-olds looked longer when the non-preferred object was reached for. Cesana-Arlotti and Halberda (2022) also found that 2.5-year-olds, who fail the 4-cup task, nonetheless reason by exclusion across word-learning, social-learning, and explicit negation with a common saccade pattern: they saccade to the to-be-excluded item, return to the target item, and fail to show "redundant" saccades—evidence of low-confidence—after target selection. This pattern suggests a domain-general inferential mechanism that delivers high-confidence conclusions, a functional profile one should expect if children perform DS.

Leahy and Carey (2020) provide an alternative, non-DS-based explanation of successful reasoning by exclusion *via* sequentially simulating alternative possibilities. However, chimpanzees, at least, are able to represent distinct possible states of affairs simultaneously. Engelmann et al. (2021) used a modified two-cup task in which the empty cup was not revealed. Chimps could pull ropes for both cups, or pull just one rope for one cup, causing the second cup to fall out of reach. Overwhelmingly they expended extra energy to pull both ropes when the cups were opaque, but pulled just one when the cups were transparent (Fig. 5b).[9] Pulling two ropes is hedging under uncertainty, suggesting chimps simultaneously represent two locations as possibly reward-laden.

Furthermore, 12-month-olds seem to use the same computations adults do to reason by exclusion, as measured by oculomotor signatures (Cesana-Arlotti et al. 2018). It's possible that adults do *both* DS and simulation-based or icon-based reasoning in these tasks. But given independent reasons to think these tasks run on LoT-like object representations in VWM and adults' capacity for DS, and the relative lack of evidence for multiple redundant reasoning processes underlying task performance, our working hypothesis is that infants's oculomotor behavior is evidence for LoT-based DS.

[9] Chimpanzees, orangutans, monkeys and children under four fail to hedge in this way when rewards are dropped in a transparent Y-shaped tube: they place a hand under just one of the arms at the bottom (Redshaw & Suddendorf 2016; Suddendorf et al. 2017; Suddendorf et al. 2019; Lambert et al. 2018). It is plausible that participants rely on simulation (Leahy & Carey 2020) here. Unlike the cups task, the Y-tube task requires anticipating the trajectory of an object that is both plainly visible and already in motion, which might encourage simulation.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Logical inference without language is a rapidly developing research area, and central contributors to this research such as Carey are skeptical of the "thicker" interpretations of the data we defend. While we anticipate further plot twists will emerge in study of infant and non-human inference, we take the current state of the literature to favor a LoT-based account of DS in infants and animals and to bear promise for many LoTH-based lines of research in the development of logical operators.

6. LoT in Social Psychology: The Logic of System 1

One source of opposition to LoTH stems from treatments of attitudes and system-1 processing in social psychology. In traditional dual-process theory, System 1 ("S1") is governed by shallow heuristic, associative, non-rule-based processing (Sloman 1996; Evans & Stanovic 2013). Dual-process theories originate partly from the heuristics-and-biases tradition, where fast responding purportedly demonstrates irrationality (cf. Gigerenzer and Gaissmaier 2011; Mandelbaum 2020b).

One may doubt the irrationality of S1 processing. As case studies we'll discuss two paradigms used to investigate characteristically S1 thought: unconscious reasoning in implicit attitudes in the Implicit Association Test and Belief Bias cases (though the same morals hold for other paradigms such as Base Rate Inferences and Cognitive Reflection Test: De Neys & Glumicic, 2008; De Neys & Franssens, 2009; Thompson, Turner, & Pennycook 2011; Stupple, Ball, Evans, & Kamal-Smith, 2011; De Neys, Cromheeke, & Osman 2011; De Neys, Rossi, & Houdé 2013; Pennycook et al. 2014; Thompson & Johnson, 2014; Gangemi, Bourgeois-Gironde, & Mancini, 2015; Johnson, Tubau, & De Neys 2016; Bago & De Neys 2017, 2019, 2020).[10]

6.1 Logic, Load, and LoT

Failures of syllogistic reasoning are commonplace and well-publicized. In particular, belief biases—cases where people mistakenly utilize the truth of a conclusion in judging an argument's validity, ignoring logical form—are legion (Markovitz & Nantel 1989). Even outside of the belief

[10] One reason S1 is so instructive is that its operations occur outside working memory. Cognition that is most plausibly governed by internal rehearsal of natural language or "inner speech" plausibly requires verbal-working-memory resources (Baddeley 1992; Marvel & Desmond 2012; Carruthers 2018). Evidence of LoT-like structure in S1 therefore undermines attempts to reduce LoT-like effects to inner speech.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

bias people are forever affirming the consequent, denying the antecedent, and confusing validity and truth.

Difficulties in reasoning are prima facie problematic for LoTH. The more errors we make in reasoning, the less it seems like we need an inferential apparatus to explain people's thinking. LoT is tailor-made to explain formal reasoning—that is, reasoning based on the structure, rather than the content, of one's premises (Fodor & Pylyshyn 1988; Quilty-Dunn & Mandelbaum 2018a; 2018b). So, failures in reasoning—traditionally seen as due to heuristic S1 processing—are seen as reasons for believing that S1 is associative rather than LoT-like (see, e.g., Sloman 1996, Rydell & McConnell 2006, Gawronski & Bodenhausen 2006). However, a closer look at the data shows evidence for non-associative, LoT-like, logic-sensitive reasoning in S1.

"Conflict problems" are cases where validity and believability conflict, i.e., valid syllogisms with unbelievable conclusions or invalid syllogisms with believable conclusions. All other problems (valid/believable; invalid/unbelievable) are "nonconflict". Some examples:

(Conflict: Valid/Unbelievable)
P1: All birds fly
P2: Penguins are birds
C: Penguins fly

(Conflict: Invalid/Believable)
P1: All birds fly
P2: Penguins are birds
C: Penguins swim

(No Conflict: Valid/Believable)
P1: All birds have feathers
P2: Penguins are birds
C: Penguins have feathers

(No Conflict: Invalid/Unbelievable)
P1: All birds have feathers
P2: Penguins are birds
C: Penguins fly

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

If S1 is not logic-sensitive, then conflict problems should not hamper believability judgments, since belief bias is driven by nonlogical factors. Yet logic-sensitive judgments occur even when subjects are explicitly instructed to focus on believability, and even under extreme cognitive load. Logical responses thus seem to be generated automatically. People are less confident and slower on conflict problems than nonconflict problems regardless of whether they are judging belief or logic (Handley & Trippas, 2015; Trippas, Thompson, & Handley 2017; Howarth, Handley, & Polito 2021). That is, they'll be slower to judge that 'Penguins fly' is false if it is a conclusion of a valid argument than a conclusion of an invalid one. Moreover, those who correctly solve syllogism validity questions in conflict problems do so even under intense time pressure and additional memory load, ensuring the shutdown of system 2 processes (Bago & De Neys 2017). That is, correct responding happens right away; giving participants additional time to think adds little accuracy.

Just as the believability of a conclusion can interfere with validity judgments, so too can the logical form of an argument affect believability judgments. In fact, there is evidence that logical responding is more automatic than belief-based responding; derailing logical responding impedes belief-based responding more than vice versa (Handley, Newstead, & Trippas 2011; Howarth, Handley, & Walsh 2016; Trippas et al. 2017). For example, in Trippas et al. (2017), conflict impeded believability judgments more than validity judgments for modus ponens. Sensitivity to logical form persists whether subjects are under load or not (and whether asked to evaluate validity or not), showing that the relevant differences are due to S1 processing (Trippas, Handley, Verde, & Morsanyi 2016). Even when asked to respond randomly, participants still show implicit sensitivity to logical form (Howarth et al. 2021). Automatic logical sensitivity also has very little individual difference between subjects, suggesting it reflects fundamental architectural features of cognition (Ghasemi, Handley, & Howarth 2021). Logical inferences are also made automatically during reading (Lea 1995; Lea, Mulligan, & Walton 2005; Dabkowski & Feiman 2021). As one would expect if logic was intuitive, subliminally presented premises trigger modus ponens inferences (Reverberi, Pischedda, Burigo, & Cherubini 2012).

Far from undermining LoTH, dual-process architectures vindicate LoTH. They demonstrate abstract logic-based inferential promiscuity outside controlled, conscious cognition using discrete symbols that maintain role-filler independence (e.g., P must be the same symbol in P—>Q).

6.2 The Logic of Implicit Attitudes

Implicit attitudes are typically assumed to be associative. However, Mandelbaum (2016) and De Houwer (2019) documented the effects of "logical interventions" on implicit attitudes, i.e., cases

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

where one can change implicit attitudes not by counterconditioning or extinction, as would be expected if they had associative structure, but instead by merely changing the logically pertinent evidence. Logical (or "propositional") interventions on attitudes are only possible given that we have predicate-argument structure, logical operators, and inferential promiscuity.

Take Kurdi and Dunham (2021). Their basic paradigm consisted of a learning and testing phase. In a learning phase participants saw sentences of the form: "If you see a green circle, you can conclude that Ibbonif is trustworthy; if you see a purple pentagon, you can conclude that Ibbonif is malicious." This design cleverly pits associative vs. propositional (i.e. LoT) processes against each other: if the implicit attitude processor is associative then Ibboniff should come out as neutral as Ibboniff is being associated with both positive (trustworthy) and negative (malicious) adjectives. If the processor is sensitive to propositional values however, then the implicit attitude acquired should be dependent on which conditional's antecedent was satisfied (i.e., which shape appears). Participants then moved onto the testing phase which consisted of explicit and implicit attitude testing (via the IAT). Results showed that participant attitudes tracked the logical form of the stimuli during the testing phase. So, using the sample text above, if participants saw a purple pentagon they would conclude that Ibbonif (and the group that he was from, the Niffites, denoted from the suffix on the name) was negatively valenced.

Kurdi & Dunham had ample variations on the paradigm all showing similar LoT-based effects on implicit attitudes. Importantly, LoT-based inferences can be seen *even when the response is normatively inappropriate*, as in an affirming-the-consequent syllogism (study 3). In the learning phase, participants saw sentences such as "If you see a green circle, you can conclude that Ibbonif is malicious;" however, instead of seeing a green circle, they would then see an (e.g.,) orange square. Thus the correct inference to make is that nothing can be inferred from the set-up. If implicit attitudes are updated only by an associative processor, then the valence of the predicate in the consequent should dictate the participants' responses. If instead attitudes are sensitive to the logical form of the inventions, then one of two things should happen: for those subjects who correctly realize that this is an affirming the consequent argument they should form no opinion about the person or group in question. However, the subset of people who incorrectly affirm the consequent should make the wrong inference and infer that the consequent accurately describes the person or group in question. Participants were given a control question to see if they were apt to explicitly affirm the consequent. Those that did also changed their implicit attitudes in line with the affirming-the-consequent stimuli they would later see in the experiment; the implicit attitudes of those who rejected the affirming the consequent control question, on the other hand, correctly tracked the logical implications of the stimuli by failing to update at all (similar results hold for denying the antecedent). Given a sufficiently creative set up, one can

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

infer logical processes at play even in the *absence* of inference, or during misinference (Quilty-Dunn & Mandelbaum 2018a).

Similar variations abound. If the associative account were correct then merely giving a major premise that is clearly valenced should set the associative value of the target: giving participants sentences such as “If you see a purple pentagon, you can conclude that Ibbonif is malicious' should make one associate IBBONIF and negative valence via ‘malicious.’ Except that isn’t what happens—if subjects are given the conditional premise with no follow-up they withhold forming any valenced implicit attitudes, unlike what associative theory would predict.[11] The concept IBBONIF needs to be linked with the attribute MALICIOUS in a way that is impervious to associative factors, but sensitive to counterevidence. A predicate-argument structure with MALICIOUS as predicate and IBBONIF as argument predicts just this functional profile.

The Kurdi and Dunham is just one of a near-deluge of recent studies showing the efficacy of logical interventions compared to the impotence of associative interventions (De Houwer 2006; Gast & De Houwer 2013; Van Dessel, De Houwer, Gast, Smith, & De Schryver 2016; Van Dessel, Gawronski, Smith, & De Houwer 2017a; Van Dessel, Mertens, Smith, & De Houwer 2017b; Van Dessel, Ye, & De Houwer 2019; Mann & Ferguson 2015, 2017; Cone & Ferguson 2015; Mann, Cone, Heggeseth, & Ferguson 2019). Telling participants that they will see a pairing of a group with pictures of pleasant (or unpleasant) things is much more effective at fixing implicit attitudes than repeatedly pairing the group and the pleasant/unpleasant things. One-shot learning trumps 37 associative pairings. Even when associative and one-shot propositional learning are combined, the associative trials add no detectable valence to the implicit attitude formed from the one-shot propositional trial (Kurdi & Banaji 2017). That is, direct exposure to associative pairings isn’t necessary or sufficient for forming or changing implicit attitudes, and its effect on attitudes doesn’t compare to a single exposure to a sentence. Even when repeated exposure causes some mental representation of the categories to be formed, just telling participants whether the stimuli are diagnostic modulates learning (e.g., if told the data isn’t diagnostic, learning is inhibited, and if told the data is diagnostic, learning is increased). This suggests that the representations acquired are being used as beliefs (Quilty-Dunn & Mandelbaum 2018b), and updated in a logical, inferentially promiscuous way (Kurdi & Banaji 2019). The primacy of diagnostic information over repeated exposure is a consistent finding, showing the inadequacies of associative models (e.g., Mann & Ferguson 2015, 2017; Mann et al. 2019).

[11] We don’t deny that there are associations in S1, just that they suffice to explain the data.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

In short, implicit attitudes—far from being a problem-area for LoT—instead demand evidence-sensitive, inferentially promiscuous predicate-argument structures that incorporate abstract logical operators.

7. Conclusion

More than half a century after the cognitive revolution of the 1950s, mental representations remain the central theoretical posits of psychology. While our picture of the mind has gotten more and more complex over time, computational operations over structured symbols remain foundational to our explanations of behavior. At least some of these symbols—those involved in certain aspects of probabilistic inference, concept acquisition, S1 cognition, object-based and relational perceptual processing, infant and animal reasoning, and likely elsewhere—are couched in a LoT. That doesn't mean that *all* perceptual and cognitive processing is LoT-symbol manipulation. We believe in other vehicles of thought, including associations (Quilty-Dunn & Mandelbaum 2020), icons (Quilty-Dunn 2020b), and much more. Our claim is somewhat modest: many representational formats across many cognitive systems are LoTs.

We don't deny the successes of DCNNs; perhaps they accurately model some aspects of biological cognition (Buckner 2019; Shea 2021). It remains open that DNNs might mimic the performance of biological perception and cognition across a wide variety of domains and tasks by *implementing* core features of LoTs (cp. Zhu et al. 2020). We agree with a recent review of DCNNs that a "key question for current research is how structured representations and computations may be acquired through experience and implemented in biologically plausible neural networks" (Peters & Kriegeskorte 2021, 1137). Given the evidence above, matching the *competences* of biological minds will require implementing a class of structured representations that uses discrete constituents to encode abstract contents and organizes them into inferentially promiscuous predicate-argument structures that can incorporate logical operators and exhibit role-filler independence.

There is much more to say about evidence for LoT, including abstract, compositional reasoning in aphasics (Varley 2014), and potential neural underpinnings for LoT (Wang et al. 2019; Frankland & Greene 2020; Roumi et al. 2021; Gershman 2022). LoTs ought to provide "common codes" that interface across diverse systems (Pylyshyn 1973; Dennett 1978). Central topics here include LoTs at the interfaces of language (Dunbar and Wellwood 2016; Pietroski 2018; Harris 2022) and action (Mylopoulos 2021; Shepherd 2021).

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

The big picture is that LoTH remains a thriving research program. LoTH allows us to distinguish psychological kinds in a remarkably fine-grained way, offering promising avenues for future research. LoTs might differ across systems within a single mind, or between species (Porot 2019). While it's likely, for example, that object tracking and S1 reasoning differ in the representational primitives they employ, we don't know whether or how their compositional principles differ. Similarly, we don't know how representations that guide logical inference in baboons differ from those that bees use in social learning, or that infants use in physical reasoning. Differences in conceptual repertoire or syntactic rules provide dimensions along which to type cognitive systems. Future work can focus on decrypting the specific symbols and transformation rules at work in each case, and how these symbols interface with non-LoT mental media.

One might also find subclusters of LoT-like properties. It may be that, for example, properties encoding logical operators and making abstract logical contents available for inference form a "logic" subcluster, and predicate-argument structure, role-filler independence, and abstract contents form a "predication" subcluster. In that case, LoT *qua* natural kind may be a genus of which these subclusters are species (as an analogy, consider how mental icons may be a genus-level kind with high species-level variation between, e.g., visual images and abstract mental models).

Finally, little is known about the evolutionary emergence of LoT in our ancestors or phylogenetically distant LoT-based minds. Our ignorance leaves open the possibility that, given LoTs' computational utility, very different biological minds converged on them independently. An outstanding research goal is to construct a typology of LoTs within and across species, allowing us to better understand the varieties of expressive power in naturally occurring representational systems (Mandelbaum et al. under review).

Acknowledgments

For helpful discussion and comments, we thank Zed Adams, Jake Beck, Ned Block, Luca Bonatti, Tyler Brooke-Wilson, Chiara Brozzo, Susan Carey, Wayne Christensen, Sam Clarke, Jonathan Cohen, Carl Craver, Roman Feiman, Chaz Firestone, Carolina Flores, EJ Green, Dan Harris, Zoe Jenkin, Benedek Kurdi, Kevin Lande, Manolo Martinez, Michael Murez, David Neely, Shaun Nichols, Jonathan Phillips, Henry Schiller, Nick Shea, Joshua Shepherd, Brent Strickland, Victor Verdejo, audiences at the Institute of Philosophy in London, Washington University in St Louis, University of Barcelona, Nantes University, the San Diego Winter

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Workshop in Philosophy of Perception, Cornell University, and the joint meeting of the SPP/ESPP. We are grateful to BBS's reviewers for comments which greatly improved the paper.

Funding statement

None.

Conflicts of interest statement

None.

References

Alter, A.L., & Oppenheimer, D.M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review* 13(3), 219–235.

Amalric, M., Wang, L., Pica, P., Figueira, S., Sigman, M., & Dehaene, S. (2017). The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS computational biology* 13(1), e1005273.

Ayzenberg, V., & Lourenco, S.F. (2021). One-shot category learning in human infants. PsyArXiv, doi:10.31234/osf.io/acymr.

Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559.

Bae, G., Olkkonen, M., Allred, S., & Flombaum, J. 2015. Why some colors appear more memorable than others: a model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General* 144, 744–763.

Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition* 158, 90–109.

Bago, B., & De Neys, W. (2019). The smart system 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning* 25(3), 257–299.

Bago, B., & De Neys, W. (2020). Advancing the specification of dual process models of higher cognition: a critical test of the hybrid model view. *Thinking & Reasoning* 26(1), 1–30.

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition* 10(8), 949–963.

Baker, N., & Elder, J. H. (2022). Deep learning models fail to capture the configural nature of human shape perception. *Iscience* 25(9), 104913.

Baker, N., Lu, H., Erlikhman, G., & Kellman, P.J. (2020). Local features and global shape information in object classification by deep convolutional neural networks. *Vision Research* 172, 46–61.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience* 5, 617–629.

Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature Reviews Neuroscience* 22(6), 359-371.

Barenholtz, E., & Feldman, J. (2003). Visual comparisons within and between object parts: evidence for a single-part superiority effect. *Vision Research* 43, 1655–1666.

Barenholtz, E., & Tarr, M.J. (2008). Visual judgment of similarity across shape transformations: Evidence for a compositional model of articulated objects. *Acta Psychologica* 128, 331–338.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences* 22, 577–609.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Bays, P.M., Catalao, R.F.G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision* 9(10), 1–11.

Bays, P.M., Wu, E.Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia* 49(6), 1622–1631.

Bermudez, J.L. (2003). *Thinking Without Words*. Oxford: OUP.

Berwick, R.C., & Chomsky, N. (2016). *Why Only Us: Language and Evolution*. Cambridge, MA: MIT Press.

Bickle, J. (2003). *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Dordrecht: Kluwer.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2), 115–147.

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition* 60(1), 1–29.

Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology* 44, 388–426.

Bowers, J. S., Malhotra, G., Dujmović, M., Montero, M. L., Tsvetkov, C., Biscione, V., … & Blything, R. (2022). Deep problems with neural network models of human vision. *PsyArXiv*, doi:10.31234/osf.io/5zf4s.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R.A. Wilson (Ed.), *Species: New Interdisciplinary Essays* (Cambridge, MA: MIT Press), 141–185.

Bracci, S., Mraz, J., Zeman, A., Leys, G., & de Beeck, H.O. (2021). Object-scene conceptual regularities reveal fundamental differences between biological and artificial object vision. BioRxiv, doi:10.1101/2021.08.13.456197.

Brady, T.F., Konkle, T., Alvarez, G.A., & Oliva, A. (2013). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Philosophy: General* 142(3), 791–808.

Braine, M.D.S. and D.P. O'Brien, eds., 1998. *Mental Logic*, Mahwah, NJ: Erlbaum.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., … & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Buckner, C. (2019). Deep learning: A philosophical introduction. *Philosophy Compass* 14(10), e12625.

Burge, T. (2010). Steps toward origins of propositional thought. *Disputatio* 4(29), 39–67.

Burgess, C.P., Matthey, L., Watters, N., Kabra, R., Higgins, I., Botvinick, M., & Lerchner, A. (2019). MONet: Unsupervised scene decomposition and representation. *arXiv preprint arXiv:1901.11390*.

Call, J. (2004). Inferences about the location of food in the great apes. *Journal of Comparative Psychology* 118 232–241.

Call, J. (2006). Descartes' two errors: Reason and reflection in the great apes. In S. Hurley & M. Nudds (Eds.), *Rational animals?* (Oxford: OUP), 219–234.

Camp, E. (2007). Thinking with maps. *Philosophical Perspectives* 21, 145–182.

Camp, E. (2009). Putting thoughts to work: Concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research* 78(2), 275–311.

Camp, E. (2018). Why maps are not propositional. In A. Grzankowski & M. Montague (eds.), *Non-propositional Intentionality* (Oxford: OUP), 19–45.

Carey, S. (2009). *The Origin of Concepts*. Oxford: OUP.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Carruthers, P. (2009). Invertebrate concepts confront the generality constraint (and win). In Lurz, R. (Ed.), *The Philosophy of Animal Minds* (New York: Cambridge University Press), 89–107.

Carruthers, P. (2018). The causes and contents of inner speech. In A. Vicente & P. Langland-Hassan (Eds.), *Inner Speech: New Voices* (Oxford: OUP), 31–52.

Carter, B., Jain, S., Mueller, J., & Gifford, D. (2021). Overinterpretation reveals image classification model pathologies. *arXiv preprint arXiv:2003.08907*.

Castelhano, M.S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review* 18, 890–896.

Cavanagh, P. (2021). The language of vision. *Perception* 50(3), 195–215.

Cesana-Arlotti, N, & Halberda, J. (2022). Domain-general logical inference by 2.5-year-old toddlers. PsyArXiv, doi:10.31234/osf.io/qzxkp.

Cesana-Arlotti, N., Kovács, A.M., & Téglás E. (2020) Infants recruit logic to learn about the social world. *Nature communications* 11(5999).

Cesana-Arlotti, N., Martín, A., Téglás, A., Vorobyova, L., Cetnarski, R., & Bonatti, L.L. (2018). Precursors of logical reasoning in preverbal human infants. *Science* 359, 1263–1266.

Cheney, D.L., & Seyfarth, R.M. (2008). *Baboon Metaphysics: The Evolution of a Social Mind*. Chicago: University of Chicago Press.

Cheng, C., & Kibbe, M.M, (2021). Children's use of reasoning by exclusion to track identities of occluded objects. *Proceedings of the Cognitive Science Society, Vol. 43.*

Cheyette, S., & Piantadosi, S. (2017). Knowledge transfer in a probabilistic Language Of Thought. In *CogSci*.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2017). Language architecture and its import for evolution. *Neuroscience and Biobehavioral Reviews* 81, 295–300.

Churchland, P.M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78, 67–90.

Clarke, S, and Beck, J. (Forthcoming). The number sense represents (rational) numbers. *Behavioral and Brain Sciences*, 1–57.

Clarke, S. (2019). Beyond the icon: Core cognition and the bounds of perception. *Mind & Language*, doi:10.1111/mila.12315.

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. Journal of Personality and Social Psychology, 108(1), 37.

Conwell, C., & Ullman, T. D. (2022). Testing Relational Understanding in Text-Guided Image Generation *ArXiv*, doi:10.48550/arXiv.2208.00005.

Dabkowski, M. & Feiman, R. (2021). Evidence of accurate logical reasoning in online sentence comprehension. Poster at the *Society for Philosophy and Psychology*.

Danks, D. (2014). *Unifying the Mind: Cognitive Representations as Graphical Models*. Cambridge, MA: MIT Press.

Davis, E., & Marcus, G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM* 58(9), 92–103.

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation* 37(2), 176–187.

De Houwer, J. (2019). Moving beyond system 1 and system 2. *Experimental Psychology* 66(4), 257–265.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition* 113(1), 45–61.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition* 106(3), 1248–1299.

De Neys, W., & Van Gelder, E. (2009). Logic and belief across the lifespan: the rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science* 12(1), 123–130.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PloS one* 6(1), e15954.

De Neys, W., Rossi, S. & Houdé, O. (2013). Bats, balls, and substitution sensitivity: cognitive misers are no happy fools. *Psychonomic Bulletin & Review* 20, 269–273.

Dickinson, A. (2012). Associative learning and animal cognition. *Philosophical Transactions of the Royal Society B* 367, 2733–2742.

Draschkow, D., & Võ, M.L.-H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Scientific Reports* 7(16471), 1–12.

Dunbar, E., & Wellwood, A. (2016). Addressing the 'two interface' problem: Comparatives and superlatives. Glossa: A Journal of General Linguistics, 1(1).

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General* 123, 501–517.

Edelman, S. (1999). *Representation and Recognition in Vision*. Cambridge, MA: MIT Press.

Egly, R., Driver, J., & Rafal, R.D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General* 123, 161–177.

Eliasmith, C. (2013). *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford: OUP.

Engelmann, J., Völter, C.J., O'Madagain, C., Proft, M., Haun, D.B., Rakoczy, H., Herrmann, E. (2021). Chimpanzees Consider Alternative Possibilities. *Current Biology* 31, R1-R3.

Erdogan, G., Yildirim, I., & Jacobs, R.A. (2015). From sensory signals to modality-independent conceptual representations: A probablistic language of thought approach. *PLoS Computational Biology* 11(11), e1004610.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science* 8(3), 223–241.

Feiman, R., Mody, S., & Carey, S. (2022). The development of reasoning by exclusion in infancy. *Cognitive Psychology* 135(101473). https://doi.org/10.1016/j.cogpsych.2022.101473

Fel, T., Felipe, I., Linsley, D., & Serre, T. (2022). Harmonizing the object recognition strategies of deep neural networks with humans. *arXiv preprint arXiv*:2211.04533

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014-18019.

Ferrigno, S., Huang, Y., & Cantlon, J.F. (2021). Reasoning through the disjunctive syllogism in monkeys. *Psychological Science* 32(2), 292–300.

Field, H. H. (1978). Mental representation. *Erkenntnis*, *13*(1), 9-61.

Finn, C., Yu, T., Zhang, T., Abbeel, P., & Levine, S. (2017, October). One-shot visual imitation learning via meta-learning. In *Conference on robot learning* (pp. 357-368). PMLR.

Firestone, C. (2020). Performance vs. competence in human-machine comparisons. *Proceedings of the National Academy of Sciences* 117(43), 26562–26571.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Firestone, C., & Scholl, B. J. (2014). “Please tap the shape, anywhere you like” shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological science*, *25*(2), 377-386.

Fitch, W.T. (2019). Animal cognition and the evolution of human language: why we cannot focus solely on communication. *Philosophical Transactions of the Royal Society B* 375, doi:10.1098/rstb.2019.0046.

Flombaum, J.I., Kundey, S.M., Santons, L.R., & Scholl, B.J. (2004). Dynamic object individuation in rhesus macaques: A study of the tunnel effect. *Psychological Science* 15(12), 795–800.

Flombaum, J.I., & Scholl, B.J. (2006). A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 840–853.

Flombaum, J.I., Scholl, B.J., & Santos, L.R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In B. M Hood & L.R. Santos (eds.), *The Origins of Object Knowledge*, 135–164. Oxford: Oxford University Press.

Fodor, J.A. (1975). *The Language of Thought* (Vol. 5). Harvard university press.

Fodor, J.A. (1981). The present status of the innateness controversy.

Fodor, J.A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: OUP.

Fodor, J.A. (2007). The revenge of the given. In B. McLaughlin and J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind* (Oxford: Blackwell), 105–116.

Fodor, J.A., & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* 28, 3–71.

Forster, M., Leder, H., & Ansorge, U. (2013). It felt fluent, and I liked it: subjective feeling of fluency rather than objective fluency determines liking. *Emotion* 13(2), 280-289.

Fougnie, D., & Alvarez, G.A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature–store model. *Journal of Vision* 11(12), 1–12.

Frankland, S.M., & Greene, J.D. (2020). Concepts and compositionality: in search of the brain's language of thought. *Annual Review of Psychology* 71, 273–303.

Futo, J., Teglas, E., Csibra, G., & Gergely, G. (2010). Communicative Function Demonstration induces kind-based artifact representation in preverbal infants. *Cognition* 117, 1–8.

Gallistel, C.R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.

Gallistel, C.R. (2011). Prelinguistic Thought. *Language Learning and Development* 7, 253–262.

Gangemi, A., Bourgeois-Gironde, S., & Mancini, F. (2015). Feelings of error in reasoning—in search of a phenomenon. *Thinking & Reasoning* 21(4), 383–396.

Gast, A., & De Houwer, J. (2013). The influence of extinction and counterconditioning instructions on evaluative conditioning effects. *Learning and Motivation* 44(4), 312–325.

Gauker, C. (2011). *Words and Images: An Essay on the Origin of Ideas*. Oxford: OUP.

Gautam, S., Suddendorf, T., & Redshaw, J. (2021). When can young children reason about an exclusive disjunction? A follow up to Mody and Carey (2016). *Cognition* 207, doi:10.1016/j.cognition.2020.104507

Gawronski, B., & Bodenhausen, G.V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological bulletin* 132(5), 692-731.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Gayet, S., Paffen, S., & Van der Stigchel, S. (2018). Visual working memory storage recruits sensory processing areas. *Trends in Cognitive Sciences* 22(3), 189–190.

Gershman, S. J. (2022). The molecular memory code and synaptic plasticity: a synthesis. *arXiv preprint arXiv:2209.04923*.

Ghasemi, O., Handley, S.J., & Howarth, S. (2021). The Bright Homunculus in our Head: Individual Differences in Intuitive Sensitivity to Logical Validity. *Quarterly Journal of Experimental Psychology*, 17470218211044691.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology* 62, 451–482.

Goldstone, R.L., Medin, D.L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology* 23, 222–262.

Goodman, N., Mansinghka, V., Roy, D., Bonawitz, K., & Tenenbaum, J. (2008). Church: A language for generative models. In D. McAllester & P. Myllymaki (Eds.), *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence, UAI 2008* (Corvallis, OR: AUAI Press), 220–229.

Goodman, N.D. and Lassiter, D. (2015). Probabilistic Semantics and Pragmatics Uncertainty in Language and Thought. In *The Handbook of Contemporary Semantic Theory* (eds S. Lappin and C. Fox), 655-686.

Goodman, N.D., Tenenbaum, J.B., & Gerstenberg, T. (2015). Concepts in a probabilistic language of thought. In E. Margolis and S. Laurence (Eds.), *Concepts: New Directions* (Cambridge, MA: MIT Press), 623-654.

Goodman, N.D., Tenenbaum, J.B., Feldman, J., & Griffiths, T. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, 32(1), 108–154.

Goodman, N.D., Ullman, T.D., & Tenenbaum, J.B. (2011). Learning a theory of causality. *Psychological Review* 118(1), 110-199.

Gordon, R.D., & Irwin, D.E. (1996). What's in an object file? Evidence from priming studies. *Perception and Psychophysics* 58(8), 1260–1277.

Gordon, R.D., & Irwin, D.E. (2000). The role of physical and conceptual properties in preserving object continuity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(1), 136–150.

Gordon, R.D., & Vollmer, S.D. (2010). Episodic representation of diagnostic and non-diagnostic object color. *Visual Cognition* 18(5), 728–750.

Gordon, R.D., Vollmer S.D., & Frankl M.L. (2008). Object continuity and the transsaccadic representation of form. *Perception and Psychophysics* 70, 667–679.

Green, E.J. (2019). On the perception of structure. *Noûs* 53(3), 564–592.

Green, E.J. (unpublished). A pluralist perspective on shape constancy.

Green, E.J., & Quilty-Dunn, J. (2021). What is an object file? *British Journal for the Philosophy of Science* 72(3), 665–699.

Gröndahl, T., & Asokan, N. (2022). Do Transformers use variable binding? *ArXiv* doi: 10.48550/2203.00162.

Hafri, A., & Firestone, C. (2021). The perception of relations. *Trends in Cognitive Sciences* 25(6), 475–492.

Hafri, A., Bonner, M.F., Landau, B., & Firestone, C. (2021). A phone in a basket looks like a knife in a cup: The perception of abstract relations. PsyArXiv, doi:10.31234/osf.io/jx4yg.

Handley, S.J., & Trippas, D. (2015). Dual processes and the interplay between knowledge and structure: A new parallel processing model. In *Psychology of learning and motivation* (Vol. 62, pp. 33-58). Academic Press.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Handley, S.J., Newstead, S.E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 28-43.

Harman, G. (1973). *Thought*. Princeton: Princeton University Press.

Harris, D. W. (2022). Semantics without semantic content. Mind & Language, 37(3), 304-328.

Harrison, S.A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458(7238), 632–635.

Haugeland, J. (1985). *Artificial intelligence: the very idea.* Cambridge: MIT Press

Hein, E., Stepper, M. Y., Hollingworth, A., & Moore, C. M. (2021). Visual working memory content influences correspondence processes. *Journal of Experimental Psychology: Human Perception and Performance* 47(3), 331–343.

Heinke, D., Wachman, P., van Zoest, W., & Leek, E.C. (2021). A failure to learn object shape geometry: Implications for convolutional neural networks as plausible models of biological vision. *Vision Research* 189, 81–92.

Hinzen, W., & Sheehan, M. (2013). *The Philosophy of Generative Grammar*. Oxford: OUP.

Hollingworth, A., & Franconeri, S. L. (2009). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface feature cues. *Cognition* 113(2), 150–166.

Hollingworth, A., & Rasmussen, I.P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance* 36(3), 543–564.

Howard, S.R., Avargues-Weber, A., Garcia, J.E., Greentree, A.D., & Dyer, A.G. (2018). Numerical ordering of zero in honeybees. *Science* 360, 1124–1126.

Howarth, S., Handley, S. J., & Walsh, C. (2016). The logic-bias effect: The role of effortful processing in the resolution of belief–logic conflict. *Memory & Cognition*, *44*(2), 330-349.

Howarth, S., Handley, S., & Polito, V. (2021). Uncontrolled logic: intuitive sensitivity to logical structure in random responding. *Thinking & Reasoning*, 1-36. DOI: 10.1080/13546783.2021.1934119

Hummel, J.E. (2000). Where view-based theories break down: The role of structure in shape perception and object recognition. In E. Dietrich and A. Markman (Eds.), *Cognitive Dynamics: Conceptual Change in Humans and Machines* (Hillsdale, NJ: Erlbaum), 157–185.

Hummel, J.E. (2011). Getting symbols out of a neural architecture. *Connection Science* 23(2), 109–118.

Hummel, J.E. (2013). Object recognition. In D. Reisburg (Ed.), *Oxford Handbook of Cognitive Psychology* (Oxford: OUP), 32–46.

Hutto, D.D. & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, MA: MIT Press.

Jiang, H. (2020). Effects of transient and nontransient changes of surface feature on object correspondence. *Perception* 49(4), 452–467.

Johnson, E.D., Tubau, E., & De Neys, W. (2016). The doubting system 1: Evidence for automatic substitution sensitivity. *Acta psychologica* 164, 56-64.

Johnson-Laird, P. (2006). *How We Reason*. Oxford: OUP.

Jordan, K.E., Clark, K., & Mitroff, S.M. (2010). See an object, hear an object file: Object correspondence transcends sensory modality. *Visual Cognition* 18(4), 492–503.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Kahneman, D., Treisman, A., & Gibbs, B.J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology* 24(2), 175–219.

Kaiser, D., Quek, G.L., Cichy, R.M., & Peelen, M.V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences* 23(8), 672–685.

Katz, Y., Goodman, N.D., Kersting, K., Kemp, C., & Tenenbaum, J.B. (2008). Modeling semantic cognition as logical dimensionality reduction. *Proceedings of the cognitive science society (Vol. 30)*.

Kelemen, D., & Carey, S. (2007). The essence of artifacts: Developing the design stance. In E. Margolis & S. Laurence (Eds.), *Creations of the Mind: Theories of Artifacts and Their Representation* (Oxford: OUP), 212–230.

Kemp, C. (2012). Exploring the conceptual universe. *Psychological Review, 119*(4), 685–722.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences* 105(31), 10687–10692.

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology* 10(11), e1003915.

Kibbe, M.M., & Leslie, A.M. (2011). What do infants remember when they forget? Location and identity in 6-month-olds' memory for objects. *Psychological Science* 22(12), 1500–1505.

Kibbe, M.M., & Leslie, A.M. (2019). Conceptually rich, perceptually sparse: Object representations in 6-month-old infants' working memory. *Psychological Science* 30(3), 362–375.

Kosiorek, A.R., Sabour, S., Teh, Y.W., & Hinton, G.E. (2019). Stacked capsule autoencoders. *arXiv preprint arXiv:1906.06818*.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The Case for Mental Imagery*. Oxford: OUP.

Kosslyn, S.M., (1980). *Image and Mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S.M., Ball, T.M., & Reiser, B.J. (1978). Visual images preserve metric spatial information: Evidence from studies of imagery scanning. *Journal of Experimental Psychology: Human Perception and Performance* 4, 47–60.

Kriegeskorte, N. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science* 2015 1:1, 417-446.

Kulvicki, J. (2015). Maps, pictures, and predication. *Ergo* 2(7), doi:10.3998/ergo.12405314.0002.007.

Kurdi, B., & Banaji, M.R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes?. *Journal of Experimental Psychology: General* 146(2), 194.

Kurdi, B., & Banaji, M.R. (2019). Attitude change via repeated evaluative pairings versus evaluative statements: Shared and unique features. *Journal of Personality and Social Psychology* 116(5), 681.

Kurdi, B., & Dunham, Y. (2021). Sensitivity of implicit evaluations to accurate and erroneous propositional inferences. *Cognition 214*, 104792.

Lake, B.M., Ullman, T.D., Tenenbaum, J.B., & Gershman, S.J. (2017). Building machines that learn and think like people. Behavioral and brain sciences, 40.

Lambert, M.L. and Osvath, M. (2018). Comparing chimpanzees' preparatory responses to known and unknown future outcomes. *Biology Letters* 14(9). https://doi.org/10.1098/rsbl.2018.0499

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Lea, R.B. (1995). On-line evidence for elaborative logical inferences in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21(6), 1469-1482.

Lea, R.B., Mulligan, E.J., & Walton, J.L. (2005). Accessing distant premise information: How memory feeds reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(3), 387-395.

Leahy, B. & Carey, S. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences* 24(1), 65–78.

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

Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing what and where systems. *Trends in Cognitive Sciences* 2(1), 10–18.

Liang, P., Jordan, M., & Klein, D. (2010). Learning programs: A hierarchical Bayesian approach. *Proceedings of the 27th International Conference on Machine Learning*, 639–646.

Lieder, F., & Griffiths, T.L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences* 1-60.

Lin, Y., Li, J., Gertner, Y., Ng, W., Fisher, C.L., & Baillargeon, R. (2021). How do the object-file and physical-reasoning systems interact? Evidence from priming effects with object arrays or novel labels. *Cognitive Psychology* 125, doi:10.1016/j.cogpsych.2020.101368.

Lonnqvist, B., Bornet, A., Doerig, A., & Herzog, M.H. (2021). A comparative biology approach to DNN modeling of vision: A focus on differences, not similarities. *Journal of Vision* 21(10), 1–10.

Loukola, O. J., Perry, C. J., Coscos, L. & Chittka, L. (2017). Bumblebees show cognitive flexibility by improving on an observed complex behavior. *Science* 355, 833-836 (2017).

Lovett, A., & Franconeri, S.L. (2017). Topological relations between objects are categorically coded. *Psychological Science* 28(10), 1408–1418.

Machery, E. (2016). The amodal brain and the offloading hypothesis. *Psychonomic Bulletin & Review* 23, 1090–1095.

Mandelbaum, E. (2013). Numerical architecture. *Topics in cognitive science*, *5*(2), 367-386.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs* 50(3), 629–658.

Mandelbaum, E. (2018). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research* 97(2), 267–283.

Mandelbaum, E. (2020a). Associationist theories of thought. *Stanford Encyclopedia of Philosophy*.

Mandelbaum, E. (2020b). Assimilation and control: belief at the lowest levels. *Philosophical Studies* 177(2), 441-447.

Mandelbaum, E., Dunham, Y., Feiman, R., Firestone, C., Green, E.J., Harris, D.W., … & Porot, N., Quilty-Dunn, J. (under review). Problems and mysteries of the many languages of thought.

Mann, T.C., & Ferguson, M.J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology* 108(6), 823-849.

Mann, T.C., & Ferguson, M.J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. *Journal of Experimental Social Psychology* 68, 122–127.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Mann, T.C., Cone, J., Heggeseth, B., & Ferguson, M.J. (2019). Updating implicit impressions: New evidence on intentionality and the affect misattribution procedure. *Journal of Personality and Social Psychology* 116(3), 349-374.

Marcus, G. F. (2001). *The Algebraic Mind*. Cambridge, MA: MIT Press.

Marcus, G. F. (2018). Deep learning: A critical appraisal. arXiv, doi:1801.00631.

Markov, Y.A., Tiurina, N.A., & Utochkin, I.S. (2019). Different features are stored independently in visual working memory but mediated by object-based representations. *Acta Psychologica*, 197, 52–63.

Markov, Y.A., Utochkin, I.S., & Brady, T.F. (2021). Real-world objects are not stored in holistic representations in visual working memory. *Journal of Vision* 21(3), 1–24.

Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition* 17(1), 11–17.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London Series B Biological Sciences* 200(1140), 269–294.

Marvel, C. L., & Desmond, J. E. (2012). From storage to manipulation: how the neural correlates of verbal working memory reflect varying demands on inner speech. *Brain and Language* 120(1), 42–51.

Martin, A.E., & Doumas, L.A.A. (2020). Tensors and compositionality in neural systems. *Philosophical Transactions of the Royal Society B* 375, doi:10.1098/rstb.2019.0306.

Meck, W.H., & Church, R.M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes* 9(3), 320-334.

Miller, J., Naderi, S., Mullinax, C., & Phillips, J. L. (2022). Attention is not enough. *Proceedings of the Annual Meeting of the Cognitive Science Society* 44(44).

Mitroff, S.R., Scholl, B.J., & Wynn, K. (2005). The relationship between object files and conscious perception. *Cognition* 96, 67–92.

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition* 154, 40–48.

Mollica, F., & Piantadosi, S. (2015). Towards semantically rich and recursive word learning models. In *Proceedings of the Cognitive Science Conference (Vol. 37)*.

Moore, C.M., Stephens, T., & Hein, E. (2010). Features, as well as space and time, guide object persistence. *Psychonomic Bulletin & Review*, *17*(5), 731–736.

Morgan, L.C. (1894). *An Introduction to Comparative Psychology.* London: Walter Scott.

Morsanyi, K., & Handley, S.J. (2012). Logic feels so good—I like it! Evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 596–616.

Mylopoulos, M. (2021). The modularity of the motor system. *Philosophical Explorations* 24(3), 376–393.

Nichols, S. (2021). *Rational Rules: Towards a Theory of Moral Learning*. New York: OUP.

Nonaka, S., Majima, K., Aoki, S.C., & Kamitani, Y. (2021). Brain hierarchy score: Which deep neural networks are hierarchically brain-like? *iScience* 24, 103013.

O'Callaghan, C. (forthcoming). Crossmodal identification. In A. Mroczko-Wąsowicz & R. Grush (Eds.), *Sensory Individuals, Properties, and Perceptual Objects* (Oxford: OUP).

Oaksford, M., & Chater, N. (2009). Précis of Bayesian rationality: The probabilistic approach to human reasoning. *Behavioral and Brain Sciences* 32(1), 69–84.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Öhlschläger, S., & Võ, M.L.-H. (2020). Development of scene knowledge: Evidence from explicit and implicit scene knowledge measures. *Journal of Experimental Child Psychology* 194(104782), 1–21.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241.

Overlan, M.C., Jacobs, R.A., & Piantadosi, S.T. (2017). Learning abstract visual concepts via probabilistic program induction in a Language of Thought. *Cognition* 168, 320–334.

Palangi, H., Smolensky, P., He, X., & Deng, L. (2018). Question-answering with grammatically-interpretable representations. *The Thirty-Second AAAI Conference on Artificial Intelligence*.

Papineau, D. (2003). Human minds. *Royal Institute of Philosophy Supplements* 53, 159–183.

Penn, D.C., Holyoak, K.J., & Povinelli, D.J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences* 31, 109–130.

Pennycook, G., Trippas, D., Handley, S.J., & Thompson, V.A. (2014). Base rates: both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40(2), 544-554.

Pepperberg, I.M., Gray, S.L., Cornero, F.M., Mody, S., & Carey, S. (2019). Logical reasoning by a Grey parrot (*Psittacus erithacus*)? A case study of the disjunctive syllogism. *Behaviour* 156, 409–445.

Perner, J., & Leahy, B. (2016). Mental files in development: Dual naming, false belief, identity and intensionality. *Review of Philosophy and Psychology* 7, 491–508.

Peters, B., & Kriegeskorte, N. (2021). Capturing the objects of vision with neural networks. ArXiv, doi:2109.03351.

Piantadosi, S.T., & Jacobs, R.A. (2016). Four problems solved by the probabilistic language of thought. *Current Directions in Psychological Science* 25(1), 54–59.

Piantadosi, S.T., Tenenbaum, J.B., & Goodman, N.D. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition* 123(2), 199–217.

Piantadosi, S.T., Tenenbaum, J.B., & Goodman, N.D. (2016). The logical primitives of thought: Empirical foundations for compositional cognitive models. *Psychological Review* 123(4), 392-424.

Pietroski, P. M. (2018). Conjoining meanings: Semantics without truth values. Oxford University Press.

Pinker, S. (1994). *The Language Instinct.* New York: William Morrow & Co.

Pollatsek, A., Rayner, K., & Collins, W.E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General* 113(3), 426–442.

Pomiechowska, B., & Gliga, T. (2021). Nonverbal category knowledge limits the amount of information encoded in object representations: EEG evidence from 12-month-old infants. *Royal Society Open Science* 8(200782), 1–17.

Porot, N.J. (2019). *Some Non-Human Languages of Thought.* Ph.D. Dissertation, CUNY Graduate Center.

Porot, N.J. (under review). Some evidence of languages of thought in chimpanzees, baboons, and an African grey parrot.

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *PNAS* 104(35), 13861–13867.

Prinz, J.J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Pylyshyn, Z.W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences* 25, 157–238.

Pylyshyn, Z.W. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.

Pylyshyn, Z.W. (2004). Some puzzling findings in multiple-object tracking: I. Tracking without keeping track of object identities. *Visual Cognition* 11(7), 801–822.

Pylyshyn, Z.W. (2007). *Things and Places: How the Mind Connects with the World.* Cambridge, MA: MIT Press.

Pylyshyn, Z.W., & Storm, R. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision* 3(3), 179–197.

Quilty-Dunn, J. (2020a). Concepts and predication from perception to cognition. *Philosophical Issues* 30(1), 273–292.

Quilty-Dunn, J. (2020b). Is iconic memory iconic? *Philosophy & Phenomenological Research* 101(3), 660–682.

Quilty-Dunn, J. (2020c). Perceptual pluralism. *Noûs* 54(4), 807–838.

Quilty-Dunn, J. (2021). Polysemy and thought: Toward a generative theory of concepts. *Mind & Language* 36, 158–185.

Quilty-Dunn, J., & Green, E.J. (forthcoming). Perceptual attribution and perceptual reference. *Philosophy and Phenomenological Research*.

Quilty-Dunn, J., & Mandelbaum, E. (2018a). Inferential transitions. *Australasian Journal of Philosophy*, *96*(3), 532–547.

Quilty-Dunn, J., & Mandelbaum, E. (2018b). Against dispositionalism: Belief in cognitive science. *Philosophical Studies* 175(9), 2353–2372.

Quilty-Dunn, J., & Mandelbaum, E. (2020). Non-inferential transitions: imagery and association In T. Chan & A. Nes, (Eds.), *Inference and Consciousness* (London: Routledge), 151–171.

Quiroga, R.Q. (2020). No pattern separation in the human hippocampus. *Trends in Cognitive Sciences* 24(12), 994–1007.

Recanati, F. (2012). *Mental Files*. Oxford: OUP.

Redshaw, J., & Suddendorf, T. (2016) Children's and apes' preparatory responses to two mutually exclusive possibilities. *Current Biology* 26, 1758–1762

Rescorla, M. (2009). Cognitive maps and the language of thought. *British Journal for the Philosophy of Science* 60(2), 377–407.

Reverberi, C., Pischedda, D., Burigo, M., & Cherubini, P. (2012). Deduction without awareness. *Acta Psychologica* 139(1), 244–253.

Richard, A.M., Luck, S.J., & Hollingworth, A. (2008). Establishing object correspondence across eye movements: Flexible use of spatiotemporal and surface feature information. *Cognition* 109(1), 66–88.

Rips, L.J. (1994). *The Psychology of Proof.* Cambridge, MA: MIT Press.

Rivera-Aparicio, J., Yu, Q., & Firestone, C. (2021). Hi-def memories of lo-def scenes. *Psychonomic Bulletin & Review* 28, 928–936.

Romano, S., Salles, A., Amalric, M., Dehaene, S., Sigman, M., & Figueira, S. (2018). Bayesian validation of grammar productions for the language of thought. *PloS One* 13(7), e0200420.

Roumi, F.A., Marti, S., Wang, L., Amalric, M., & Dehaene, S. (2021). Mental compression of spatial sequences in human working memory using numerical and geometrical primitives. *Neuron* 109, 2627–2639.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Rumelhart, D.E., & McClelland, J.L. (1986). *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations*. Cambridge, MA: MIT Press.

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. (3rd ed.). Prentice Hall.

Rydell, R.J., & McConnell, A.R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology* 91(6), 995–1008.

Sablé-Meyer, M., Ellis, K., Tenenbaum, J., & Dehaene, S. (2021a). A language of thought for the mental representation of geometric shapes. PsyArXiv, doi:10.31234/osf.io/28mg4.

Sablé-Meyer, M., Fagot, J., Caparos, S., van Kerkoerle, T., Amalric, M., & Dehaene, S. (2021b). Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity. *Proceedings of the National Academy of Sciences*, *118*(16), e2023123118.

Saiki, J., & Hummel, J.E. (1998a). Connectedness and part-relation integration in shape category learning. *Memory & Cognition* 26(6), 1138–1156.

Saiki, J., & Hummel, J.E. (1998b). Connectedness and the integration of parts with relations in shape perception. *Journal of Experimental Psychology: Human Perception and Performance* 24(1), 227–251.

Schneider, S. (2011). *The Language of Thought: A New Philosophical Direction*. Cambridge, MA: MIT Press.

Scholl, B.J. (2007). Object persistence in philosophy and psychology. *Mind & Language* 22(5), 563–591.

Scholl, B.J., & Leslie, A. (1999). Explaining the infant's object concept: Beyond the perception/ cognition dichotomy. In E. Lepore & Z. W. Pylyshyn (Eds.), *What Is Cognitive Science?* (Oxford: Blackwell), 26–73.

Scholl, B.J., Pylyshyn, Z.W., & Franconeri, S.L. (Unpublished). The relationship between property–encoding and object–based attention: Evidence from multiple object tracking.

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., ... & DiCarlo, J. J. (2018). Brain-score: Which artificial neural network for object recognition is most brain-like? *BioRxiv*, doi:10.1101/407007.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., ... & Silver, D. (2020). Mastering atari, go, chess and shogi by planning with a learned model. *Nature* 588(7839), 604–609.

Schwitzgebel, E. (2013). A dispositional approach to attitudes: Thinking outside of the belief box. In N. Nottelman (ed.), *New Essays on Belief: Constitution, Content, and Structure* (New York: Palgrave Macmillan), 75–99.

Shea, N. (2018). *Representation in Cognitive Science*. Oxford: OUP.

Shea, N. (2021). Moving beyond content-specific computation in artificial neural networks. *Mind & Language*, doi:10.1111/mila.12387.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171, 701–703.

Shepherd, J. (2021). Intelligent action guidance and the use of mixed representational formats. *Synthese* 198(17), 4143–4162.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin* 119(1), 3-22.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46, 159–216.

Smortchkova, J., & Murez, M. (forthcoming). Representational kinds. In J. Smortchkova, K. Dolega, & T. Schlicht (eds.), *What are Mental Representations?* Oxford: OUP.

Solvi, C., Al-Khudhairy, S.G., & Chittka, L. (2020). Bumble bees display cross-modal object recognition between visual and tactile senses. *Science* 367, 910–912.

Spelke, E.S. (1990). Principles of object perception. *Cognitive Science* 14, 29–56.

Stavans, M., & Baillargeon, R. (2018). Four-month-old infants individuate and track simple tools following functional demonstrations. *Developmental Science* 21, e12500.

Stavans, M., Lin, Y., Wu, D., & Baillargeon, R. (2019). Catastrophic individuation failures in infancy: A new model and predictions. *Psychological Review* 126(2), 196–225.

Stein, T., Kaiser, D., & Peelen, M.V. (2015). Interobject grouping facilitates visual awareness. *Journal of Vision* 15(8), 1–11.

Strickland, B., & Scholl, B.J. (2015). Visual perception involves event-type representations: The case of containment versus occlusion. *Journal of Experimental Psychology: General* 144(3), 570–580.

Stupple, E.J., Ball, L.J., Evans, J.S.B., & Kamal-Smith, E. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology* 23(8), 931–941.

Suddendorf, T., Crimston J., & Redshaw, J. (2017). Preparatory responses to socially determined, mutually exclusive possibilities in chimpanzees and children. *Biology Letters*, 13(6). doi.org/10.1098/rsbl.2017.0170

Suddendorf, T., Watson, K., Bogaart, & M., Redshaw, J. (2019). Preparation for certain and uncertain future outcomes in young children and three species of monkey. *Developmental Psychobiology* 62(2), pp. 191-201.

Surian, L., & Caldi, S. (2010). Infants' individuation of agents and inert objects. *Developmental Science* 13(1), 143–150.

Szabo, Z. (2011). The case for compositionality. In W. Hinzen, E. Machery & M. Werning (Eds.), *The Oxford Handbook on Compositionality* (Oxford: OUP), 64–80.

Tenenbaum, J.B., Kemp, C., Griffiths, T.L., & Goodman, N.D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science* 331(6022), 1279–1285.

Thompson, V.A., & Johnson, S.C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning* 20(2), 215–244.

Thompson, V.A., Turner, J.A.P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology* 63(3), 107–140.

Tikhonenko, P.A., Brady, T.F., & Utochkin, I.S. (2021). Independent storage of real-world object features is visual rather than verbal in nature. PsyArXiv, doi:10.31234/osf.io/d9c4h.

Tolman, E.C. (1948). Cognitive maps in rats and men. *Psychological* Review, 55(4), 189–208.

Toribio, J. (2011). Compositionality, iconicity, and perceptual nonconceptualism. *Philosophical Psychology* 24(2), 177–193.

Travis, C. (2001). *Unshadowed Thought*. Cambridge, MA: Harvard University Press.

Trippas, D., Handley, S.J., Verde, M.F., & Morsanyi, K. (2016). Logic brightens my day: Evidence for implicit sensitivity to logical validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(9), 1448-1457.

Trippas, D., Thompson, V.A., & Handley, S.J. (2017). When fast logic meets slow belief: Evidence for a parallel-processing model of belief bias. *Memory & Cognition* 45(4), 539–552.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Tuli, S., Dasgupta, I., Grant, E., & Griffiths, T. L. (2021). Are Convolutional Neural Networks or Transformers more like human vision? ArXiv, doi:2105.07197.

Ullman, S. (1996). *High-level Vision*. Cambridge, MA: MIT Press.

Ullman, T.D., Goodman, N.D., & Tenenbaum, J.B. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development* 27(4), 455–480.

Utochkin, I.S., & Brady, T.F. (2020). Independent storage of different features of real-world objects in long-term memory. *Journal of Experimental Psychology: General* 149(3), 530–549.

Van Dessel, P., De Houwer, J., Gast, A., Smith, C.T., & De Schryver, M. (2016). Instructing implicit processes: When instructions to approach or avoid influence implicit but not explicit evaluation. *Journal of Experimental Social Psychology* 63, 1–9.

Van Dessel, P., Gawronski, B., Smith, C.T., & De Houwer, J. (2017a). Mechanisms underlying approach-avoidance instruction effects on implicit evaluation: Results of a preregistered adversarial collaboration. *Journal of Experimental Social Psychology* 69, 23–32.

Van Dessel, P., Mertens, G., Smith, C.T., & De Houwer, J. (2017b). The mere exposure instruction effect. *Experimental psychology 64(*5):299-314.

Van Dessel, P., Ye, Y., & De Houwer, J. (2019). Changing deep-rooted implicit evaluation in the blink of an eye: Negative verbal information shifts automatic liking of Gandhi. *Social Psychological and Personality Science* 10(2), 266–273.

Varley, R. (2014). Reason without much language. *Language Sciences* 46, 232–244.

Vasas, V., & Chittka, L. (2019). Insect-inspired sequential inspection strategy enables an artificial network of four neurons to estimate numerosity. *iScience* 11, 85–92.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Võ, M.L.-H. (2021). The meaning and structure of scenes. *Vision Research* 181, 10–20.

Võ, M.L.-H., & Henderson, J.M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision* 9(3), 1–15.

Võ, M.L.-H., & Wolfe, J.M. (2013). Different electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science* 24(9), 1816–1823.

Võ, M.L.-H., Bettcher, S.E.P., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology* 29, 205–210.

Wang, B., Cao, X., Theeuwes, J., Olivers, C.N.L., & Wang, Z. (2017). Separate capacities for storing different features in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(2), 226–236.

Wang, L., Amalric, M., Fang, W., Jiang, X., Pallier, C., Figueira, S., ... & Dehaene, S. (2019). Representation of spatial sequences using nested rules in human prefrontal cortex. *NeuroImage*, 186, 245–255.

Webb, T. W., Sinha, I., & Cohen, J. D. (2020). Emergent symbols through binding in external memory. *ArXiv*, doi:10.48550/arXiv.2012.14601

Weise, C., Ortiz, C. C., & Tibbetts, E. A. (2022). Paper wasps form abstract concept of 'same and different.' *Proceedings of the Royal Society B: Biological Sciences*, *289*(1979), 20221156. https://doi.org/10.1098/rspb.2022.1156

Wood, J.N., & Wood, S.M.W. (2020). One-shot learning of view-invariant object representations in newborn chicks. *Cognition* 199, doi:10.1016/j.cognition.2020.104192.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press

Xu, F. (2019). Toward a rational constructivist theory of cognitive development. *Psychological Review* 126(6), 841–864.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology* 30(2), 111–153.

Xu, Y. (2017). Reevaluating the sensory account of visual working memory storage. *Trends in Cognitive Sciences* 21(10), 794–815.

Xu, Y. (2020). Revisit once more the sensory storage account of visual working memory. *Visual Cognition* 5-8, 433–446.

Xu, Y., & Vaziri-Pashkam, M. (2021a). Examining the coding strength of object identity and nonidentity features in human occipito-temporal cortex and convolutional neural networks. *Journal of Neuroscience* 41(19)*,* 4234–4252.

Xu, Y., & Vaziri-Pashkam, M. (2021b). Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nature Communications* 12(2065), 1–16.

Xu, Y., Zhou, X., Chen, S., & Li, F. (2019). Deep learning for multiple-object tracking: A survey. *IET Computer Vision* 13(4), 355–368.

Yamins, D.L., & DiCarlo, J.J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience* 19(3), 356–365.

Yassa, M.A., & Stark, C.E.L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences* 34(10), 515–525.

Ye, X., & Durrett, G. (2022). The unreliability of explanations in few-shot in-context learning. *ArXiv*, doi:2205.03401

Yildirim, I., & Jacobs, R.A. (2015). Learning multisensory representations for auditory-visual transfer of sequence category knowledge: a probabilistic language of thought approach. *Psychonomic Bulletin & Review* 22(3), 673–686.

Zhou, K., Luo, H., Zhou, T., Zhuo, Y., & Chen, L. (2010). Topological change disturbs object continuity in attentive tracking. *PNAS* 107(50), 21920–21924.

Zettlemoyer, L.S., & Collins, M. (2005). Learning to map sentences to logical form: Structured classification with probabilistic categorical grammars. *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, 658–666.

Zhu, Y., Gao, T., Fan, L., Huang, S., Edmonds, M., Liu, H., ... & Zhu, S.C. (2020). Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense. *Engineering* 6(3), 310–345.

https://doi.org/10.1017/S0140525X22002849 Published online by Cambridge University Press