of the six core properties are observed in each of the few cognitive domains discussed in the paper: Three properties are demonstrated by implicit social cognition and four (the maximal number of cooccurring core properties) in the object-files case. Shall we conclude that only a few cognitive domains involve language-like representations? An interesting conclusion surely, but one that is much less exciting than the one touted by Quilty-Dunn et al.

Second, Quilty-Dunn et al. haven't even shown that clustering of the core properties is a unique prediction of the language-of-thought hypothesis. Many of the core properties are in fact coinstantiated in neural networks. For instance, the outputs of a sequence-to-sequence language model like BERT evince (at the very least) role-filler independence and predicate–argument structure (in addition to the general capacity for abstraction demonstrated by neural networks). Evidence suggests that these characteristics are underlain by systematic syntactic and semantic competences (Clark, Khandelwal, Levy, & Manning, 2019; Tenney, Das, & Pavlick, 2019). Thus, other architectures are consistent with the clustering of the core properties.

Perhaps the authors think that the burden-of-proof is on their opponents to show that these other formats exist and can account for the apparent clustering. But outside philosophy, such burden-of-proof claims are as weak an argument as it gets. Inferring a language-of-thought architecture on such shaky grounds also runs the risk of slowing research in computational neuroscience on new alternative cognitive architectures that are both neuroscientifically plausible and that can account for the core properties. Finally, and most important, alternatives to language-of-thought cognitive architectures have been investigated for decades, and the properties discussed by Quilty-Dunn et al. are known to result from these (Eliasmith, 2013; Eliasmith & Anderson, 2003; Smolensky, 1990, 1991). In none of these cases do the architectures merely implement a language-of-thought.

Finally, Quilty-Dunn et al. rely on the epistemic virtues of explanatory breadth and unification to support the language-of-thought hypothesis: As they say, "The chief aim […] is to showcase LoTH's explanatory breadth and power in light of recent developments in cognitive science" (target article, sect. 1, para. 3). But an appeal to explanatory breadth runs against their pluralistic commitment: If the authors are serious about representational pluralism, it is hard to understand why they believe that explanatory breadth is a virtue or why any unification should be expected.

Although their defense of the language-of-thought hypothesis fails, Quilty-Dunn et al. are onto something important: We should expect cognition to exploit the core properties to solve some types of cognitive challenges, and we should thus predict their occurrence in some cognitive domains. Which tasks are facilitated by these properties and which life forms in the phylogenetic tree had to solve such tasks (and why) are exciting empirical questions.

**Competing interest.** None.

## References

Bansal, K., Loos, S., Rabe, M., Szegedy, C., & Wilcox, S. (2019). *HOList: An environment for machine learning of higher order logic theorem proving*. Proceedings of the 36th international conference on machine learning (Vol. 97, pp. 454–463).

Clark, A. (1993). *Associative engines: Connectionism, concepts, and representational change*. MIT Press.

Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). *What does BERT look at? An analysis of BERT's attention*. Proceedings of the 2019 ACL workshop BlackboxNLP: Analyzing and interpreting neural networks for NLP (pp. 276–286). Florence, Italy: Association for Computational Linguistics.

Dai, W. Z., Xu, Q., Yu, Y., & Zhou, Z. H. (2019). *Bridging machine learning and logical reasoning by abductive learning*. Proceedings of the 33rd international conference on neural information processing systems (pp. 2811–2822). Red Hook, NY: Curran Associates Inc.

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. MIT Press.

Irving, G., Szegedy, C., Alemi, A. A., Eén, N., Chollet, F., & Urban, J. (2016). DeepMath-deep sequence models for premise selection. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 29, pp. 2235–2243).

Machery, E. (2014). In defense of reverse inference. *The British Journal for the Philosophy of Science*, *65*, 251–267.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data?. *Trends in Cognitive Sciences*, *10*(2), 59–63.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, *46*(1–2), 159–216.

Smolensky, P. (1991). Connectionism, constituency, and the language of thought. In B. M. Loewer & G. Rey (Eds.), *Meaning in mind: Fodor and his critics* (pp. 201–227). Oxford: Blackwell.

Stoianov, I., & Zorzi, M. (2012). Emergence of a "visual number sense" in hierarchical generative models. *Nature Neuroscience*, *15*(2), 194–196.

Tenney, I., Das, D., & Pavlick, E. (2019). *BERT rediscovers the classical NLP pipeline*. Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 4593–4601). Florence, Italy: Association for Computational Linguistics.

# Perception is iconic, perceptual working memory is discursive

Ned Block

Department of Philosophy, New York University, New York, NY, USA
Ned.block@nyu.edu
https://www.nedblock.us

doi:10.1017/S0140525X23001899, e265

**Abstract**

The evidence that the target article cites for language-of-thought (LoT) structure in perceptual object representations concerns perceptual working memory, not perception. Perception is iconic, not structured like an LoT. Perceptual working memory representations contain the remnants of iconic perceptual representations, often recoded, in a discursive envelope.

In their wonderful and provocative target article, Quilty-Dunn et al. say perceptual object representations have language-of-thought (LoT) structure. However, there is plenty of evidence that perceptual object representations are iconic in a sense that excludes LoT representations; the evidence Quilty-Dunn et al. cite pertains to discursive *perceptual working memory* (WM) representations, not discursive *perceptual representations*. I will first present some evidence that perceptual object representations are iconic, then that WM representations are discursive. I will use the term "discursive" for representations that exhibit almost all of the six properties they cite rather than "LoT" because I doubt that even perceptual working memory exhibits all of them. (See Susan Carey's response to the target article.) But if there are no discursive representations in perception, there are no LoT representations either.

Apparent motion suggests iconic perceptual object representations. When two nearby objects flicker with the right parameters,

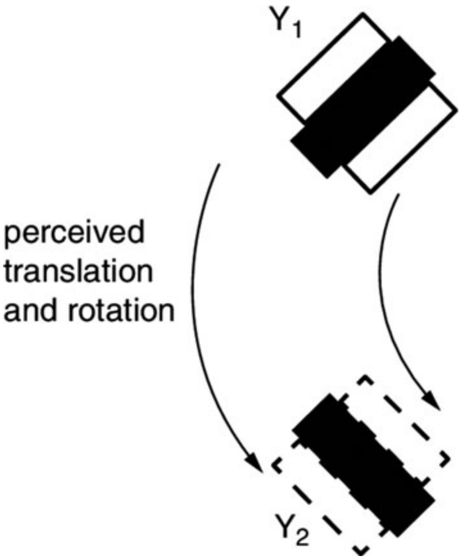

**Figure 1** (Block). Items at Y1 and Y2 flicker so as to create apparent motion between them. The shapes are viewed in an apparatus in which a slightly different image is projected to each eye. This allows one version in which the black bar is part of a squarish shape and another version in which the bar protrudes, making the item look like an object instead of a shape. In the latter case, the subject sees rotation along with the movement but in the former case the subject sees just motion. Thanks to Ken Nakayama for providing this figure.

we see motion between them. Objects move while visible properties change gradually.

See the caption to Figure 1. What suggests iconicity in this case of apparent motion is analog mirroring: Certain relations in the world are mirrored by representations that instantiate analogs of those relations in a way that is sensitive to degrees of difference. What is interesting about this case is that when the figure is perceived as an object, mirroring respects objecthood.

If the flickering objects are of different sizes, we see smooth expansion and contraction. One might suppose that the further apart the flickering objects are, the faster the rate of flicker would have to be to see motion. However, mirroring dictates the opposite because objects that are further apart take longer to traverse the distance. The further apart the flickering objects are, the longer the time span between flickers has to be to see motion (Korte's Third Law). The visual system prefers short motion paths between flickering objects but that preference is overridden if the shortest path involves biologically impossible motion (Shiffrar & Freyd, 1990) or if a moving object turns into an object of a different kind. In sum, perceptual representations are iconic in a sense that excludes discursive representations (see Block, 2023a, 2023b).

I now switch to the topic of perceptual working memory. Perceptual working memory often contains the remnants of perception – typically *not consciously experienced*. It can include iconic materials, but visual working memory often includes them in recoded form. One illustration of the partially nonperceptual nature of visual working memory is illustrated in Figure 2. When the central disk and the donut surrounding it are presented simultaneously, there is center-surround suppression on the right, but not the left. However, when they are presented one at a time, with the first stimulus maintained in working memory, the collinear effect disappears (Bloem, Watanabe, Kibbe, & Ling, 2018). Thus a fundamental computational aspect of perception is absent in this working memory representation (see pp. 113–114 of Block, 2023a).

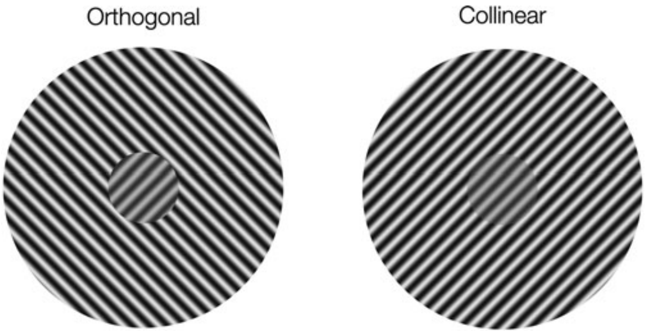

**Figure 2** (Block). Central disk is the same on the left and on the right but it looks much higher in contrast on the left. The suppressive center/surround effect in the collinear case is because of a fundamental computation known as divisive normalization. Thanks to Sam Ling for providing this figure.

Quilty-Dunn et al.'s arguments for the iconic nature of perception involve the "object files" of perceptual working memory. But working memory representations in visual areas are often recoded outside of the classic visual system, for example, in the intraparietal sulcus, while they disappear from visual cortex because of ongoing visual stimulation (Rademaker, Chunharas, & Serences, 2019). Perceptual information is often recoded in the service of specific tasks. Kwak and Curtis (2022) showed subject clouds of moving dots and also oriented gratings, asking them to remember the directions. They found that brain decoding on either of these working memory representation worked on the other suggesting that working memory coded what was in common to the two kinds of percepts, eliminating the moving dots and the gratings, replacing them with representations of vectors, showing that many iconic features can be altered or discarded in working memory.

Of course perceptual working memory is constantly interacting with perception. Quilty-Dunn (2023) argues convincingly that this interaction is crucial to longer term perceptions, for example, perceptions that span saccades. As Quilty-Dunn notes, perception does not start anew after each saccade, so there must be some perceptual – I would say iconic – information preserved by the saccade. True, but there is plenty of evidence for at least some loss of iconicity in transsaccadic memory (summarized in Block, 2023a, pp. 261–262). For example, in the famous Sperling effect, a multirow array of letters is presented briefly, but a cue presented after the stimulus stops can focus attention on any one of the rows, allowing reporting of all or almost all the items. However, if the array is presented before the saccade and the cue presented afterward, the Sperling effect disappears, showing that transsaccadic memory can erase the iconic memory that the Sperling effect depends on.

Quilty-Dunn describes a perceptual effect known as the motion repulsion illusion. If dots moving in one direction are superimposed on dots moving in a different direction, the perceived angle between the two directions is exaggerated in perception. Kang et al. showed that the same effect occurs if one set of moving dots is seen while the other is held in working memory (Kang, Hong, Blake, & Woodman, 2011). This result suggests that there are iconic elements in perceptual working memory but does nothing to show that perception is not iconic.

As Quilty-Dunn notes, perception can distort working memory and conversely. Teng and Kravitz (2019) showed that colors and orientations in each of perception and working memory affect the other, commenting that this is no doubt

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

because of overlapping representations in visual processing areas. However, as the authors note, these results are compatible with the involvement of prefrontal cortex in working memory. The overlap of sensory coding between visual working memory and vision does not preclude partial reformatting in working memory or the inclusion of iconic information in a discursive envelope.

In sum, perceptual working memory can preserve some aspects of iconic perceptual representation even if it includes it in a discursive envelope. Quilty-Dunn et al.'s results depend on the discursive envelope, not the iconic perceptual representations.

**Financial support.** The author acknowledges Templeton World Charities for support.

**Competing interest.** None.

## References

Block, N. (2023a). *The border between seeing and thinking*. Oxford University Press. Open access at https://global.oup.com/academic/product/the-border-between-seeing-and-thinking-9780197622223?cc=us&lang=en&

Block, N. (2023b). Let's get rid of the concept of an object file. In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 494–516). Wiley Blackwell.

Bloem, I. M., Watanabe, Y. L., Kibbe, M. M., & Ling, S. (2018). Visual memories bypass normalization. *Psychological Science*, *29*(5), 845–856. doi:10.1177/0956797617747091

Kang, M.-S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, *18*(5), 860–869. doi:10.3758/s13423-011-0126-5

Kwak, Y., & Curtis, C. E. (2022). Unveiling the abstract format of mnemonic representations. *Neuron*, *110*(11), 1822–1828. doi:https://doi.org/10.1016/j.neuron.2022.03.016

Quilty-Dunn, J. (2023). *Remnants of perception: Comments on Block*. Paper presented at the American Philosophical Association, San Francisco, April 6, 2023.

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, *22* (8), 1336–1344. doi:10.1038/s41593-019-0428-x

Shiffrar, M., & Freyd, J. J. (1990). Apparent motion of the human body. *Psychological Science*, *1*(4), 257–264. doi:10.1111/j.1467-9280.1990.tb00210.x

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour*, *3*(8), 827–836. doi:10.1038/s41562-019-0640-4

# Natural logic and baby LoTH

Irene Canudas-Grabolosa[a], Ana Martín-Salguero[b,c] and Luca L. Bonatti[b,d]

[a]Department of Psychology, Department of Linguistics, Harvard University, Cambridge, MA, USA; [b]Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain; [c]Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Saclay, NeuroSpin Center, Gif/Yvette, France and [d]ICREA, Barcelona, Spain
irenecanudas@gmail.com, ana.martin@upf.edu, lucabonatti@mac.com

doi:10.1017/S0140525X23001942, e266

**Abstract**

Language-of-thought hypothesis (LoTH) is having a profound impact on cognition studies. However, much remains unknown about its basic primitives and generative operations. Infant studies are fundamental, but methodologically very challenging. By distilling potential primitives from work in natural-language semantics, an approach beyond the corset of standard formal logic may be undertaken. Still, the road ahead is challenging and long.

Fodor had the gift of conceiving extremely simple ideas with extremely deep and rich consequences. Language-of-thought hypothesis (LoTH) is perhaps the best, but not the unique, example of this gift. Quilty-Dunn et al.'s article is a very forceful testimony of how lively and far-reaching LoTH is. The very fact that they use a cluster of properties that prescinds from most traditional arguments for LoT is in itself a proof of its richness. At the same time, as it is clear in the target article, like other cases (modularity witness it), Fodor's LoTH was more a research program than an hypothesis; in his words, it's probably a genus, but, we would add, one whose actual species are still barely known. This dearth of knowledge is particularly acute for one of the fundamental issues in characterizing LoT(s): Identify the basic primitives available endogenously in human thinking. In adults, a recent work investigated modular LoTs defined over various domains, proposing primitives and compositional routines (Al Roumi, Marti, Wang, Amalric, & Dehaene, 2021; Dehaene, Al Roumi, Lakretz, Planton, & Sablé-Meyer, 2022; Planton et al., 2021; Sablé-Meyer et al., 2021; Sablé-Meyer, Ellis, Tenenbaum, & Dehaene, 2022). However exciting and important to characterize human singularity, these theories do not clarify the origins of LoT or its role in general human cognition. They can check out all the list of properties in Quilty-Dunn et al.'s cluster, and yet remain confined to the specific domain they have been tested, in adults. They are compatible with the fact that language interactions, or instruction, contributes to their appearance.

Although there is little doubt that when linguistic competence kicks in, human language competence is explained by reference to a system of structures encompassing many properties of a general LoT, the crucial open questions are whether properties of general thinking are somehow imported from linguistic structures, as many would hold (Carruthers, 2002; Spelke, 2003), or else are inherent properties of the mind, and if they encompass the logical concepts that make LoT cross-domain and compositional. Progress on these questions can be achieved by investigating the existence and nature of the logical primitives available to preverbal infants, who are likely not affected by instructions or massive language experience. Unfortunately, these investigations can be counted on the fingers of one hand. They are also very difficult, as they require creating scenes deprived of verbal cues that likely embed logical inferences, something that at best can be supported by arguments to the best explanation. We thank Quilty-Dunn et al., who agree with us that a baby-LoTH has the upper hand relative to alternative theories, but they are more optimistic than us: Alternative explanations, perhaps compatible, perhaps incompatible with some declination of LoTH (Leahy & Carey, 2020), exist and have to be addressed experimentally. Furthermore, an unified explanation of the early putative indications of logical thinking (Cesana-Arlotti et al., 2018; Cesana-Arlotti, Kovács, & Téglás, 2020; Cesana-Arlotti, Téglás, & Bonatti, 2012; Cesana-Arlotti, Varga, & Téglás, 2022) and the later failures at making action plans consistent with it (Feiman, Mody, & Carey, 2022; Leahy, Huemer, Steele, Alderete, & Carey, 2022; Mody & Carey, 2016) is still missing. All these issues require painstaking research.

As baby LoTH supporters, we believe that the most serious question remains the identification of a plausible repertoire of early LoT primitives. Short of the success in disjunctive reasoning (Cesana-Arlotti et al., 2018), little exists about other logical components of an LoT, while some arguably plausible candidates – for