Perceptual (roots of) core knowledge

Brian J. Scholl* *Corresponding author.
Department of Psychology, Yale University, New Haven, CT, USA brian.scholl@yale.edu https://perception.yale.edu/
doi:10.1017/S0140525X23003023, e140

Abstract
Some core knowledge may be rooted in – or even identical to – well-characterized mechanisms of mid-level visual perception and attention. In the decades since it was first proposed, this possibility has inspired (and has been supported by) several discoveries in both infant cognition and adult perception, but it also faces several challenges. To what degree does What Babies Know reflect how babies see and attend?

Introduction: What babies see?
As the various subfields of cognitive science have become ever more distinct and specialized, the notion of core knowledge has acted as a sort of intellectual glue – synergizing research from its intellectual origins in developmental psychology, to studies of animal cognition, adult visual perception, linguistic representation, computational modeling and AI, and beyond. Here I focus on one particular form of synergy, between What Babies Know (Elizabeth Spelke’s brilliant and groundbreaking book summarizing one of the most productive research programs in all of science; Spelke, 2022; henceforth WBK) and the study of what and how we see (as explored in studies of adult visual perception and attention).

What kinds of mental representations and processes characterize core knowledge? Once upon a time, the answer was unambiguous: Higher-level thought. As Spelke once suggested, “Humans come to know about an object’s unity, boundaries, and persistence in ways like those by which we come to know about its material composition or its market value” (Spelke, 1988, p. 198). This view was inspired by a (now-obsolete) characterization of perception as relatively unsophisticated. Chapter 2 of WBK, for example, involves discrete objects, but as Spelke once suggested: “Perceptual systems do not package the world into units… The parsing of the world into things may point to the essence of thought and to its essential distinction from perception” (1988, p. 229). And since infants continue to represent objects that are not currently in view, the responsible mechanisms must therefore “carry infants beyond the world of immediate perception” (p. 172). By the 1990s, however, advances in the study of adult perception had made it clear that visual processing does in fact “package the world into units” on its own, independent of higher-level thought – into representations of both surfaces (for an early review, see Nakayama, He, & Shimojo, 1995) and objects (for an early review, see Scholl, 2001), which then persist through time, occlusion, and featural change (for a review, see Scholl, 2007).

These discoveries led to a proposal, first articulated in the late 1990s and early 2000s (Leslie, Xu, Tremoulet, & Scholl, 1998; Scholl, 2001, Section 7.2; Scholl & Leslie, 1999; see also Carey & Xu, 2001) that at least some types of core knowledge may be rooted in the mechanisms and representations of mid-level visual processing and object-based attention (henceforth mid-level vision). Infants may have expectations about the behaviors of objects not because of considered deliberation or conceptual theories, but because that is simply how they experience the world in the first place, in terms of their brute visual percepts. “[S]urprising parallels between recent results in cognitive developmental psychology and the study of object-based visuospatial attention suggest that the two areas of inquiry may have something to do with each other” (Scholl & Leslie, 1999, p. 60) – and although “visual processing in adults may seem relatively unrelated to the study of core knowledge in infant cognition, … recent work has suggested that these two seemingly different fields may in fact be studying the same underlying representations and constraints” (Strickland & Scholl, 2015, p. 571).

Progress: Sophisticated seeing!
The ultimate value of any theoretical proposal lies in the concrete progress it inspires. How has the proposal that core knowledge is rooted in mid-level vision fared in the decades since it was first introduced? Here are three examples of how this view has fueled new discoveries in both domains:

Cohesion and persistence: Many core knowledge principles apply to objects but not non-solid substances, and early work showed how cohesion violations (failures to maintain rigid boundaries and internal connectedness) frustrate infants’ object tracking (e.g., Huntley-Fenner, Carey, & Solimando, 2002; Spelke & Van de Walle, 1993). This led directly to the discoveries in adult vision that cohesion violations also frustrate attentional tracking (vanMarle & Scholl, 2003) and the maintenance of object-file representations – even when just viewing a single object split into two (Mitroff, Scholl, & Wynn, 2004). And this adult vision work then directly inspired the demonstration that even a single object (e.g., a cracker) splitting into two destroys infants’ ability to track quantity (Cheries, Mitroff, Wynn, & Scholl, 2008).

Attentional prioritization: Categorizing a stimulus into a particular “event type” (such as occlusion or containment) biases infants to remember features that are especially diagnostic for that type (such as the width of an object, with a vertical container; e.g., Hespos & Baillargeon, 2001; Wang, Baillargeon, & Brueckner, 2004). This led directly to the discovery that such
prioritization also occurs spontaneously in adults’ visual working memory: While viewing dynamic containment (but not occlusion) events, change detection is better for those changes in that affect whether objects will “fit” (Strickland & Scholl, 2015) – and the subtle details of this were then subsequently also seen in infants’ object tracking (Goldman & Wang, 2019).

Seeing agency: Chapter 7 of WBK reviews many studies showing how infants automatically treat certain motion patterns (e.g., involving pursuit) as cues to agency and intentionality – and how they expect agents to behave rationally, for example by following direct paths (e.g., Gergeley, Nadasdy, Csibra, & Biro, 1995; Southgate & Csibra, 2009). This led directly to the discovery that adults’ mid-level vision also spontaneously (and even irrespectibly) extracts properties such as agency and goal-directedness when viewing “chasing” displays (Gao, Newman, & Scholl, 2009, 2010; van Buren, Uddlenberg, & Scholl, 2016) – and that violations of rational action similarly destroy adults’ ability to spontaneously see chasing (Gao & Scholl, 2011).

These examples demonstrate how taking connections between infant cognition and adult perception seriously can drive empirical progress – showing how these two domains employ similar representations (e.g., of agency), are constrained by similar principles (e.g., of cohesion), and have similar downstream consequences (e.g., of orienting attention). At the least, such parallels suggest that one domain may help to fuel the other – that core knowledge may be rooted in, and partially grow out of, mid-level vision. At the most extreme, such connections suggest that these two domains could be one and the same.

**Challenges: Prosociality, language, and beyond**

The essence of the progress reviewed above is a striking match between the results of experiments in infant cognition and adult perception. And such matches may go far beyond these three case studies (Bai, 2023), extending even into the nuances and mechanics of habituation itself (Turk-Browne, Scholl, & Chun, 2008). But just how close is this match? The biggest challenges to the view sketched above may lie in cases where the match is imperfect, in either direction.

The suggestion that core knowledge in infancy transcends mid-level vision in adults seems especially salient for at least two domains, each of which is the focus of a key chapter of WBK. First, as reviewed in Chapter 8, young infants may already have expectations and preferences related to prosociality – as when they observe one shape help (or hinder) another shape from climbing an incline (Hamilin, Wynn, & Bloom, 2007). But no work has yet suggested that visual processing itself directly extracts representations of helping, hinderin, or prosociality in general (even though properties such as [im]morality in certain visual scenes may be correlated with lower-level cues; De Freitas & Alvarez, 2018). Second, as reviewed in Chapter 9, several aspects of core knowledge seem intimately related to language. Even infants’ object tracking, for example, can depend on how people linguistically refer to the objects (Dewar & Xu, 2007; Xu, 2002). But mid-level vision seems largely encapsulated from linguistic processing, and vice versa (Firestone & Scholl, 2016) – and so if core knowledge reflects the operation of mid-level vision, then such linguistic connections may be rendered mysterious or inexplicable.

Potential mismatches may also loom large in the other direction – when adults’ mid-level vision seems more sophisticated than infants’ core knowledge. The studies reviewed in the previous section are all examples in which visual representations have been found to be especially sophisticated – encompassing properties and constraints (such as agency and cohesion) more closely associated with higher-level thought. But this trend in vision research goes far beyond the classical domains of core knowledge.

Additional work, for example, has suggested that mid-level vision automatically and spontaneously extracts representations of causal history (i.e., of how objects came to look the way that they do; Chen & Scholl, 2016), soft-material intuitive physics (e.g., inferring the shape of objects under cloths; Wong, Bi, Soltani, Yildirim, & Scholl, 2023), and even unfinishedness (as when an object appears not to have ended its movement; Ongchoco, Wong, & Scholl, 2023). But competence involving such seemingly sophisticated domains is nowhere to be found in most characterizations of core knowledge.

On one hand, some of these challenges could be dissolved with further research. After all, when the current proposal was first articulated in the late 1990s, nobody yet suspected that mid-level vision might match infant cognition in the ways reviewed in the previous section. And so we might still discover that mid-level vision extracts representations of prosociality, or that infant core knowledge also encompasses representations of causal history. On the other hand, some of these challenges remain despite having been recognized long ago (e.g., Scholl & Leslie, 1999, Section 5.5) – and without a principled way to demarcate which results we expect to “match” and which we do not (e.g., only spatiotemporal processing, but not contact-mechanical processing; Cheries, Mitroff, Wynn, & Scholl, 2009; Scholl & Leslie, 1999), these challenges continue to be acute.

**The state of the art**

The view of core knowledge sketched here contrasts in some ways with that from WBK. On one hand, Spelke notes that “I believe there is truth to this view” that “the object representations [involved in core knowledge] are the products of perceptual processes” (p. 78) – and throughout the book she masterfully reviews relevant work on adults’ mid-level vision (including much of the work discussed here). She also notes that her views on these issues have changed over time: “I once proposed, wrongly, that objects are not grasped by a perceptual system but by … a system of central cognition….. Research … provided decisive evidence against this proposal…. Adults were found to share the representational system found in infants” (precis, sect. 1). As such, the view sketched here is meant as more of a friendly extension than a criticism – perhaps just placing a sharper focus on certain themes from WBK.

On the other hand, WBK also provides several additional arguments against this view, which seem less compelling. Spelke suggests that some aspects of core knowledge cannot have perceptual origins because (a) core knowledge representations are abstract (p. xxi) – but many aspects of mid-level vision also abstract over many surface variables (Scholl, 2007); (b) perception involves “detectable surfaces” rather than “the entities that those surfaces belong to” (p. 198) – but recent work in mid-level vision argues for exactly the opposite view (Wong et al., 2023); and (c) core knowledge representations have a time-course that transcends momentary perception (p. 78) – but at least some object-file representations in mid-level vision have been shown to persist throughout interruptions for at least 8 seconds, and possibly much longer (a result that was also explicitly motivated by connections to infant cognition; Noles, Scholl, & Mitroff, 2005).
Some of the most recent discussions of core knowledge also still seem to veer away from the possibility of substantive connections to mid-level vision in other ways. In WBK, for example, Spelke champions the notion of the infant mind as implementing a type of “physics engine” (Ullman, Spelke, Battaglia, & Tenenbaum, 2017) – but such frameworks do not necessarily abide by the constraints of encapsulated mid-level vision, as they also readily accommodate higher-level knowledge (e.g., about how the colors of blocks may arbitrarily signal their masses; Battaglia, Hamrick, & Tenenbaum, 2013). As a result, the physics-engine framework may be true and important – but also simply orthogonal to the distinction between mid-level vision and higher-level thought. And whereas WBK ultimately characterizes core knowledge (somewhat ambiguously) as occupying “a middle ground between perceptual systems and belief systems” (p. xxi), I hope that we might also continue to take seriously the more direct possibility that core knowledge is (rooted in) mid-level vision – and that What Babies Know may largely reflect how babies see and attend.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interest. None.

References


Core knowledge and its role in explaining uniquely human cognition: Some questions

Armin W. Schulz*

Department of Philosophy, University of Kansas, Lawrence, KS, USA
awschulz@ku.edu
http://people.ku.edu/~awschulz/2023

*Corresponding author.

doi:10.1017/S0140525X23003199, e141

Abstract

Questions can be raised about the central status that evolutionarily ancient core knowledge systems are given in Spelke’s otherwise very compelling theory. So, the existence of domain-general learning capacities has to be admitted, too, and no clear reason is provided to doubt the existence of uniquely human cognitive adaptations. All of these factors should be acknowledged when explaining human thought.

https://doi.org/10.1017/S0140525X23002443 Published online by Cambridge University Press