This chapter appears (with neater figures) as: Scholl, B. J., & Leslie, A. M. (1999). In E. Lepore & Z. Pylyshyn (Eds.), *What is Cognitive Science?* Oxford: Blackwell.

2

Explaining the Infant's Object Concept: Beyond the Perception/Cognition Dichotomy

Brian J. Scholl and Alan M. Leslie

1. Introduction

Some of the most exciting research in recent cognitive science has involved the demonstration that young infants possess a remarkable array of discriminative abilities. Infants a few months old have been shown to have a substantial amount of 'initial knowledge' about objects, in domains such as physics and arithmetic (for recent reviews and overviews, see Baillargeon, 1995; Carey, 1995; Spelke, 1994; Spelke, Gutheil, & Van de Walle, 1995; Spelke, Vishton, & von Hofsten, 1995). In this chapter we will be concerned with what such results tell us about the structure of the infant's mind — in other words with what the phrase 'initial knowledge' means in terms of the underlying cognitive architecture. For the duration of this chapter, we will adopt Spelke's phrase 'initial knowledge' without scare-quotes, having recognized that it is precisely the meaning of this phrase which is at issue.

Traditional discussions have often addressed the nature of initial knowledge in terms of an implicit dichotomy between 'perception' and 'cognition' (e.g. Bogartz, Shinskey, & Speaker, 1997; Kellman, 1988; Leslie, 1988; Spelke, 1988a, 1988b). From within such a dichotomy, 'perception' has often been found want-

---

Portions of this chapter were presented by the first author at the 1997 Annual Meeting of the *Society for Philosophy and Psychology* (New York, 5/5/97). We are grateful to the following people for stimulating conversation and for helpful comments and criticism on earlier (in some cases much earlier) drafts: Richard Bogartz, Susan Carey, Jonathan Cohen, Jacob Feldman, Ron Mallon, Amy Needham, Thomas Papathomas, Zenon Pylyshyn, Richard Samuels, Elizabeth Spelke, Stephen Stich, Anne Treisman, Polly Tremoulet, Jonathan Weinberg, and Fei Xu. We also wish to thank the members of the Princeton-Rutgers-NYU Object Group for extremely helpful discussion of these issues.
Address for correspondence: Brian J. Scholl, Rutgers Center for Cognitive Science, Psychology Building Addition, Busch Campus, Rutgers University - New Brunswick, 152 Frelinghuysen Road, Piscataway, NJ, 08854-8020, USA.
Email: scholl@ruccs.rutgers.edu , **Web**: http://ruccs.rutgers.edu/~scholl/

ing as an explanation, resulting in more 'conceptual' theories which attribute to the infant various *thoughts*, *theories*, *principles*, and *beliefs* about objects. Such views are highlighted by the idea that this initial knowledge is the core which develops into and interacts with other, later-acquired beliefs in the relevant domains. At one extreme, many researchers "doubt . . . that mechanisms for apprehending objects can be distinguished from mechanisms of thought in any sense" (Spelke, 1988b, p. 220). We will refer to this as the *maximally central* view of the infant's object concept. At the other extreme, deflationary accounts of initial knowledge explicitly reject the maximally central view, and attempt to defend modified perceptual explanations (e.g. Bogartz et al., 1997; Melkman & Rabinovitch, 1998). These deflationary accounts have their origin in the traditional empiricist accounts of the origin of the object concept, whereby infants initially interact with the world via fleeting sensations, which they gradually organize into more structured entities. We will refer to theories which attempt to explain initial knowledge by appeal only to sensation as *maximally sensory* theories.

We will argue that neither the maximally central view nor the maximally sensory view is correct, and that the dichotomy between 'perceptual' and 'cognitive' explanations is of dubious value. The mechanisms and processes which drive infants' discriminative abilities may be best characterized as neither 'perceptual' nor 'conceptual', but somewhere in between. We agree with several traditional arguments that maximally sensory views cannot adequately explain the object concept. At the same time, these arguments (discussed below) do not entail a maximally central account, since there do exist mental mechanisms which are not captured by the dichotomy — for example, mechanisms of object-based visuospatial attention.

In this chapter we will explore the prospects for explaining parts of the infant's object concept by appeal to such mechanisms. Object-based mechanisms of visuospatial attention enjoy many of the crucial properties which serve to rule out maximally sensory explanations, and yet are not maximally central. We will suggest that the initial knowledge comprising the infant's object concept is best characterized in terms of this *attentional interface* between perception and cognition.[1] The attentional mechanisms at this interface may be able to account for the infants' abilities without appeal to beliefs or principles about object *types* (i.e. without appeal to general explicit beliefs about objects), but only to reactions to specific object *tokens*, via mechanisms whose existence has been motivated independently from the cognitive development literature. In this vein, we hope to draw together two literatures which have been developed completely independently: developmental research on the infant's 'object concept', and research on the nature of object-based visuospatial attention in adults (see also Leslie, Xu, Tremoulet, & Scholl, 1998).

---

[1]On visuospatial attention as the interface between perception and cognition, see Julesz (1990) and Pylyshyn (1998, in press).

In the next section we discuss the results of experiments with infants, which comprise the *explananda* for the different competing accounts. We discuss the general design of the relevant experiments, and describe in detail some examples of initial knowledge in the domains of physics and arithmetic. We then turn to the explanation of these data. In *Section 3* we identify several arguments which have been taken to favor maximally central explanations over maximally sensory explanations, and we discuss the resulting appeals to high-level thought and cognition. We then argue in *Section 4* that these arguments rest on a dubious dichotomy. We appeal to object-based mechanisms of visuospatial attention as an example of mental mechanisms that possess the crucial properties lacked by maximally sensory accounts, but that are not maximally central. The conclusion of this section will be a modest claim of possibility: that there is an alternative to both maximally sensory and maximally central accounts which rules out a strategy of arguing for one by arguing against the other.

An account of the infant's object concept which is neither maximally sensory nor maximally central may well be possible. But could such mechanisms actually explain any of the infancy data? We address this question in *Section 5*, by discussing our *Object Indexing* framework — a theory of the infant's object concept which is motivated by mechanisms at the attentional interface between perception and cognition. We discuss some related theories of the object-concept in *Section 6* (including other recent 'deflationary' explanations of the object concept, and appeals to *sortal* concepts), and we offer some concluding thoughts in *Section 7*. Throughout we use notions such as 'knowledge' without prejudging the issue of what property of infant cognitive architecture is implicated.

2. Spatiotemporally-Based Initial Knowledge in Infancy

2.1 Spatiotemporal vs. Contact-Mechanical Properties

What is the appropriate scope for a theory of the infant's object concept — i.e. for an account of what infants know about objects, where and when they must exist, and how they interact with each other? On the one hand, there has been an enormous amount of research in recent years on many kinds of initial knowledge about objects, and it might seem *a priori* implausible that a single theory could account for everything. On the other hand, it seems unfair for a theory to arbitrarily pick and choose its explananda. One wants to specify the domain of an explanation in some non-arbitrary way, which characterizes the domain as a natural kind. For the purposes of this chapter, we will draw a distinction between those parts of object knowledge which are based on *contact mechanics* and those which are not.

Some of the initial knowledge which infants seem to enjoy concerns how objects can physically interact with each other. Baillargeon, Spelke, & Wasserman (1985), for instance, investigated infants' knowledge of object *solidity*, and argued that infants are sensitive to the constraint that objects cannot occupy the same place at the same time, and thus cannot pass through each other. Other examples of such 'contact-mechanical' knowledge include Spelke's studies of *solidity and persistence* (e.g. Spelke, Breinlinger, Macomber, & Jacobson, 1992), Leslie's studies of *physical causality* (e.g. Leslie, 1984; Leslie & Keeble, 1987), and Baillargeon's studies of *support* and *collision* phenomena (summarized in Baillargeon, 1995).

At the same time, many other parts of the object concept do not seem to focus on the physical interactions of objects. Other studies ask in what conditions infants will apprehend the existence of an object in the first place, under what conditions object representations will persist, and how objects are enumerated. (We discuss several examples of this type below.) We do not intend our discussion in this chapter to speak to knowledge of contact-mechanical constraints, but only to this latter type of knowledge, of what we might call '*spatiotemporal' constraints.* We believe that there may be a basic architectural distinction between those types of initial knowledge based on contact-mechanical constraints, and those based on spatiotemporal constraints. We suggest that these two types of initial knowledge may be subserved by specific and distinct mechanisms. In any case, we address only the spatiotemporal types of initial knowledge in this chapter.

This discussion thus complements Leslie's (1994) notion of *ToBy* — the 'theory of body' mechanism — which is a more specialized (and perhaps modular) piece of core architecture on which later developing knowledge can bootstrap (cf. Leslie, 1986, 1988). Motivated by the fact that infants of certain ages seem to have initial knowledge about properties such as *solidity*, *substance*, and *causality* (e.g. Leslie & Keeble, 1987; Spelke et al., 1992), but not *inertia* or *gravity* (e.g. Kim & Spelke, 1992; Spelke et al., 1992, Experiments 4 and 5; Spelke, Katz, Purcell, & Erlich, 1994; though cf. Needham & Baillargeon, 1993), Leslie has suggested that we might think of ToBy as an embodiment of contact-mechanical knowledge — i.e. knowledge of the ways in which solid objects can (and cannot) interact with each other. See Leslie (1994) for details.

Having established the scope of our discussion, we now review the general design of the looking-time experiments which frequently characterize infant research in this area, and then discuss three demonstrations of initial 'spatiotemporal' knowledge which employ these methods.

2.2 Looking-Time Methods for Assessing Infant Knowledge

The demonstration of initial knowledge in infancy often relies on analyzing infants' *looking times*. The amount of time infants spend looking at a stimulus display decreases with repeated presentations. Several experimental paradigms based on this fact have been employed in order to determine the character of cognitive abilities in infancy.

A canonical example works as follows: When you present an infant with a stimulus, she will typically visually orient to it and focus attention on it, and then eventually look away. Repeat this event, until the infant's looking times have decreased by some specified amount. The infant is now *habituated* to (or *familiarized* with) the event. At this point, introduce some change in the stimulus event, and present it as before. If the infant detects the difference, and interprets it as a fundamentally novel stimulus, then she will *dishabituate* (or 'recover' looking time), and her looking-time will jump back up by some measurable amount. We might say that she is 'surprised', so long as we leave open what this means in terms of the infant's cognitive architecture. If she fails to detect the difference (or if she does not interpret it is a significant difference), then her looking-time will stay near the familiarized level, since she is just seeing another instance of the same familiar stimulus.

The trick, then, is to design test events and control events such that infants will look longer at those test events which violate some principle (e.g. of physics or arithmetic) than at control events which incorporate similar perceptual differences without violating the principle. By designing ingenious controls, one can work toward a precise characterization of the properties on the basis of which the infants recover looking times. If this property seems characterizable only in terms of some principle or law, then researchers can conclude that the infant has a mechanism with knowledge of that principle.

Again, though, the nature of this knowledge is exactly what is at issue here. This becomes especially clear when the results of such experiments are phrased in terms of 'surprise': "The infant was *surprised* by the 'impossible' event." Some writers (e.g. Bogartz et al., 1997) have taken issue with such descriptions, arguing that infants in such experiments do not demonstrate bona fide surprise, but merely look longer at the relevant test stimuli. As with 'initial knowledge', however, we suggest that the meaning of such phrases is exactly what is at issue.

2.3 Example #1: Spatiotemporal Continuity

Consider the following physical principle: *Objects cannot jump in and out of existence. If an object moves from point A to point B, it must do so by traversing a continuous path through space*. Do infants employ this principle, as we do, when perceiving and reasoning about the physical world? Spelke and her colleagues (Spelke & Kestenbaum, 1986; Spelke, Kestenbaum, Simons, & Wein, 1995) have demonstrated that indeed they do, based on the stimuli of Moore, Borton, and Darby (1978).

They tested four-month-old infants on displays involving an object which traveled from left to right (and back again), passing behind two occluders (see Figure 1). In the 'continuous' condition (Figure 1a), the object appeared in between the two occluders; in the 'discontinuous' condition (Figure 1b), the object did not appear between the two occluders: it disappeared behind the first, and then eventually reappeared (after an appropriate delay) from behind the second, without ever traversing the intervening space. Infants were habituated to one of these events, and were subsequently presented with a test event, consisting of the same object motions, but without the occluders. In the one-object test event (Figure 1c), a single object moved from left to right (and back again), simulating the motion of the continuous event. Likewise, the two-object test event simulated the discontinuous event (Figure 1d).

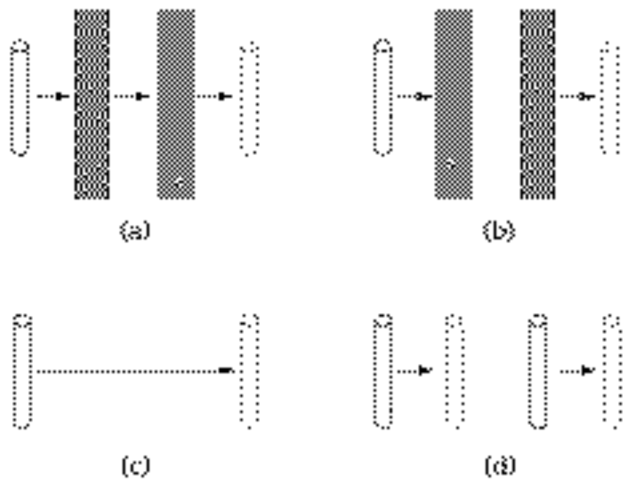

FIG. 1. Stimuli employed by Spelke, Kestenbaum, Simons, and Wein (1995): (a) the 'continuous' familiarization event; (b) the 'discontinuous' familiarization event; (c) the one-object test event; (d) the two-object test event. See text for details. (Adapted from Spelke, Kestenbaum, et al., 1995)

Looking times to these test events were compared with each other, and with control groups who saw only the test events, with no previous habituation. Infants habituated to the 'continuous' event tended to look longer at the two-object test event, while infants habituated to the 'discontinuous' event tended to look longer at the one-object test event. In other words, infants generalized more from the continuous events to single-object tests, and from the discontinuous events to two-object tests. In general: "Although preferences for the two-object display did not differ consistently from control levels over the . . . experiments, the trends in the data . . . suggest that infants perceived a single object in the continuous event and two objects in the discontinuous event" (Spelke, Kestenbaum et al., 1995, p. 136). Spelke and her colleagues (see also

Spelke, 1988a) take these results as reflecting initial knowledge of the principle of spatiotemporal continuity.[2]

2.4 Example #2: Arithmetic

Karen Wynn (1992) has reported similar experiments suggesting that infants may possess initial knowledge of numerical principles, including the arithmetical laws which define addition and subtraction. Two of the conditions from her experiment are presented in Figure 2.[3]

In her '1+1 = 1 or 2' condition (Figure 2a), five-month-old infants viewed a single doll resting on a small stage. A screen then rose up to cover part of the stage, obscuring the doll. While the screen was up, a hand appeared with a second doll, moved behind the screen, and left empty-handed. The screen then dropped, revealing either one or two dolls. Wynn measured infants' looking times to this final tableau, and found longer looking-times to the single-object case compared to the two-object case, suggesting that the infants 'expected' there to be two objects.

A separate '2–1 = 1 or 2' condition (Figure 2b) was run to be sure that infants weren't simply responding preferentially to fewer items. When two objects appear initially, and a hand removes one from behind a screen, leaving either one or two objects, infants looked longer at the two-object display. These results suggested to Wynn that either (a) infants are able to "compute the numerical results of these arithmetical operations" (Wynn, 1992, p. 750) or (b) they simply know that arithmetical operations *change* quantities of items — that addition leads to more items, and subtraction to fewer items. To test these hypotheses, she ran another '1+1 = 2 or 3' condition (not pictured), in which the result was either two or three objects. Infants looked longer at the three-item final tableaus, suggesting to Wynn that they are indeed doing arithmetic.

Simon, Hespos, & Rochat (1995) replicated Wynn's results, but added an interesting twist. Using 'Ernie' and 'Elmo' dolls as stimuli, their experiment included not only conditions with *arithmetical violations* (e.g. Ernie + Ernie = Ernie; as in Wynn, 1992) but also conditions with *identity violations* (e.g. Ernie + Ernie = Ernie + Elmo) and conditions with *both* types of violations (e.g. Ernie + Ernie = Elmo). Following Wynn (1992), the experiments tested both addition (1+1 = 1 or 2) and subtraction (2-1 = 1 or 2). (A control condition verified that the infants could indeed distinguish the two dolls.)

---

[2] Parts of these experiments were previously reported by Spelke and Kestenbaum (1986), who used slightly different test events. For another modified replication, see Xu and Carey (1996, Experiment 1).

[3] Several earlier studies have demonstrated numerical competence in infants, particularly the ability to detect correspondences between sets of items. These studies have demonstrated that infants can reliably detect differences between sets of 2 and 3 items (e.g. Antell & Keating, 1983; Starkey & Cooper, 1980; Starkey, Spelke, & Gelman, 1990; Strauss & Curtis, 1981) and sometimes between sets of 3 and 4 items (e.g. Starkey & Cooper, 1980; Von Loosbrook & Smitsman, 1990).

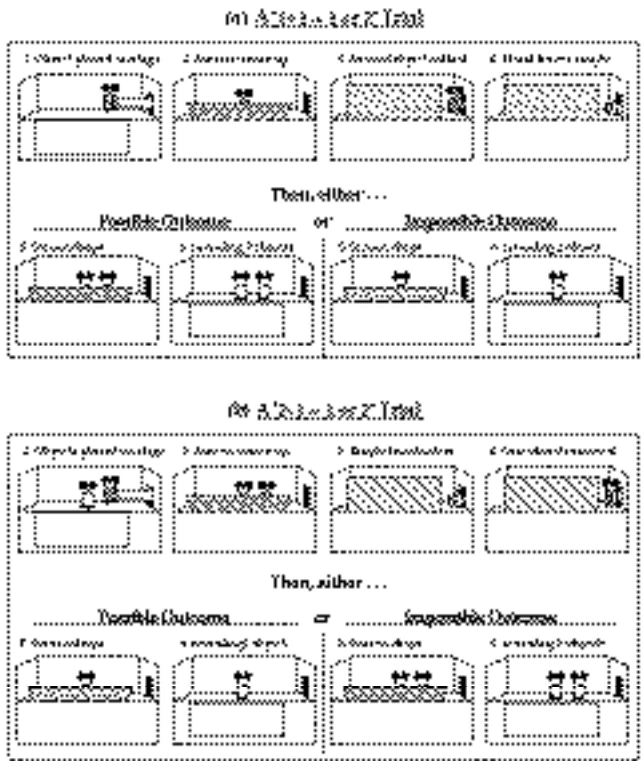

FIG. 2. Stimuli employed by Wynn (1992): (a) the '1+1 = 1 or 2' condition; (b) the '2-1 = 1 or 2' condition. See text for details. (Adapted from Wynn, 1992)

Simon et al. (1995) measured the duration of the infants' first gazes to the different outcomes, and found a tendency to look longer at arithmetical violations (compared to the final tableaus with the 'correct' arithmetical answer), but no tendency to look longer at identity violations (compared to the final tableaus containing the 'correct' dolls). These results suggest that infants expected there to be the correct number of objects in the final tableau, but didn't care about the identities of those objects, even if they had somehow (magically) changed during the period of occlusion. This experiment thus supported Wynn's contention that young infants already possess initial knowledge of at least some principles of arithmetic.[4]

---

[4] Simon et al. (1995) were actually pursuing a deflationary account of Wynn's experiments, in which infants were responding not to arithmetical violations per se, but only to the associated physical violations (i.e. of the principle that objects can't jump in and out of existence). Their

2.5 Example #3: Spatiotemporal and Property-Based Criteria for Object Apprehension

Xu and Carey (1996) have recently explored more carefully when (and on what basis) infants will apprehend the existence of an object behind a screen. In a typical experiment, infants were familiarized with events in which two objects were taken from and then replaced behind a screen. (One object was always removed and replaced from the left side of the screen, while the other was always removed and replaced from the right side of the screen.) During their removal and replacement, the two objects were either in view simultaneously (in what we will call the *spatial condition*; see Figure 3) or only sequentially (in what we'll call the *property condition*; see Figure 4). The two objects (e.g. a yellow rubber duck and a white styrofoam ball) typically differed both in their perceptual properties (i.e. size, shape, color) and in their categorical kind. Following this familiarization phase was a test phase in which the screen was removed to reveal either one or both of the previously seen objects (see Figures 3 and 4).

Ten-month old infants in the spatial condition tended to look longer at final tableaus containing only a single object, but they showed no preference for one or two test objects in the final tableaus of the property condition. Twelve-month old infants, in contrast, showed a preference for one test object in both the spatial and temporal conditions.[5]

We can interpret these results as follows. Ten-month-old infants expected there to be two objects in the final tableau of the spatial condition (and thus looked relatively longer at a single test object), but had no such expectation in the property condition. In other words, infants appeared to use *spatiotemporal* information (i.e. the simultaneous presence of both objects) as a basis on which to form an expectation for two objects, but declined to use *property/kind* information (i.e. the difference in the identities and visible properties of the two objects) as such a basis, in situations where no spatiotemporal information was available. The results of this experiment and others "are consistent with the strong claim that the property differences between the two objects had no effect at all on the [ten-month-old] baby's looking time patterns" (Xu & Carey, 1996, p. 136). This is in marked contrast, of course, to our mature perceptions of such events: adults (and 12-month-olds) will readily form an expectation for two

---

results, however, failed to support or disconfirm this hypothesis. More recently, Simon (1997, 1998) has offered another 'non-numerical' account of these results, which we discuss in Section 6.1.

[5] The data were actually more complicated than this. In all conditions there was a baseline preference for looking at two objects (compared to one), so that 'success' in these experiments consisted of overcoming this baseline preference. In the experiment described above, for instance, ten-month-old infants looked longer at a final tableau containing two objects in both the *baseline* and *property* conditions, but looked equally long at one and two test objects in the *spatial* condition (thus overcoming the baseline preference, and looking relatively longer at one test object). See Xu and Carey (1996) for details.

objects behind the screen not only when two objects are simultaneously revealed, but also if two sequentially revealed items differ in their visible properties.[6]

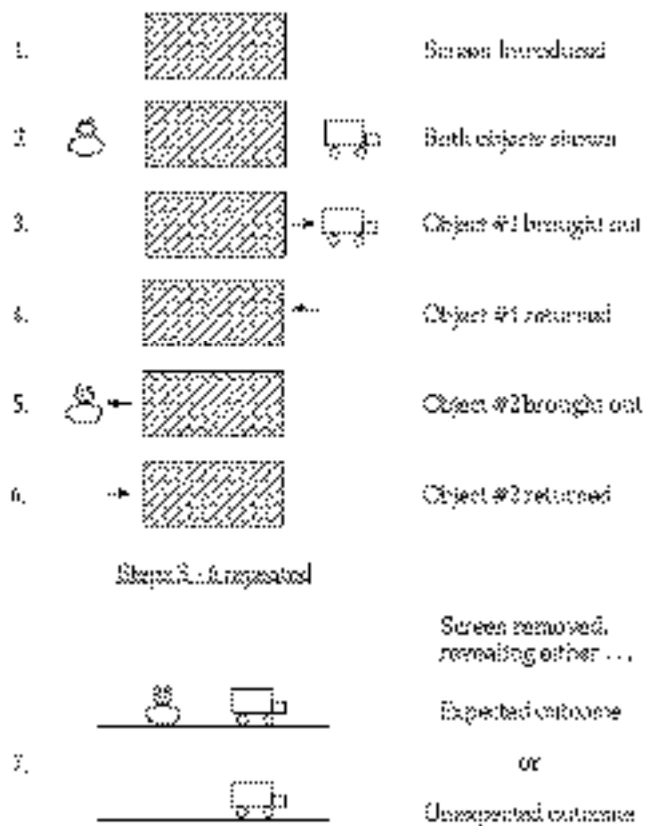

FIG. 3. The *spatial* condition from Xu and Carey (1996). See text for details. (Adapted from Xu & Carey, 1996)

In sum, these studies have demonstrated that infants possess the ability to discriminate stimuli in looking-time experiments in ways which respect various physical and arithmetical principles, in particular those principles which seem to be based on spatiotemporal factors rather than contact-mechani-

---

[6] Xu and Carey (1996) interpret their results in terms of *sortal concepts*, and more particularly as support for the *object-first hypothesis*, wherein infants employ the sortal concept BOUNDED PHYSICAL OBJECT before they employ any more specific sortal concepts. We discuss this proposal in Section 6.2.

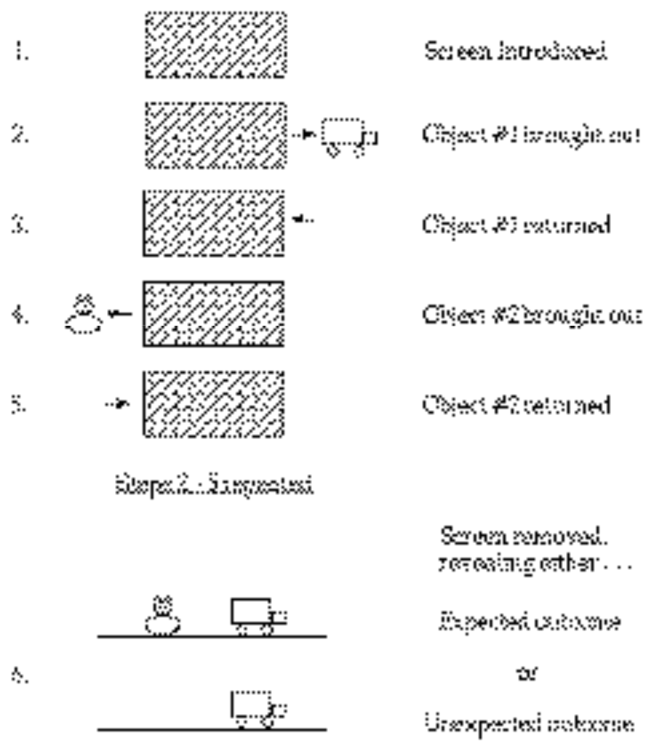

FIG. 4. The *property* condition from Xu and Carey (1996). See text for details. (Adapted from Xu & Carey, 1996)

cal factors. Xu and Carey (1996) conclude from their experiments that 10-month-old infants possess the capacity to infer the existence of an object behind a screen on the basis of earlier spatiotemporal information, but not property/kind information, while 12-month-olds can use both sorts of information when inferring the existence of occluded objects. Wynn (1992) concludes from her 'arithmetic' experiments that infants are able to "compute the numerical results of . . . arithmetical operations" (p. 750) and that "The existence of these arithmetical abilities so early in infancy suggests that humans innately possess the capacity to perform simple arithmetical calculations" (p. 750). And Spelke concludes from the continuity experiments that "Infants appear to apprehend the identity of objects by analyzing the apparent continuity or discontinuity of paths of motion, in accord with the principle that objects move on spatio-temporally continuous paths" (Spelke, 1988a, p. 179).

3. 'Conceptual' vs. 'Perceptual' Accounts of the Object Concept

Having reviewed some examples of the sorts of experiments which motivate claims of initial knowledge about objects, we now turn to the issue of exactly how this initial knowledge is embedded in the cognitive architecture. Previous discussions of the nature of infants' discriminative abilities in these domains have tended to focus on a distinction between 'perceptual' explanations and 'cognitive' or 'conceptual' explanations (e.g. Bogartz et al., 1997; Kellman, 1988; Leslie, 1988; Spelke, 1988a, 1988b). Spelke, for instance, discussed (in a paper titled 'Where perceiving ends and thinking begins') why 'perceptual' explanations are inappropriate:

> [They] assume that objects are *perceived*: that humans come to know about an object's unity, boundaries, and persistence in ways like those by which we come to know about its brightness, color, or distance. I suggest, in contrast, that objects are *conceived*: 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, 1988b, p. 198)

From within such a dichotomy between perceptual (or what we have been calling maximally sensory) explanations and conceptual (what we call maximally central) explanations, any arguments against one of these options can be taken as support for the other. And since nearly all of the relevant discussions proceeded by providing arguments that the maximally sensory explanations could not be correct, their conclusion is that the correct explanations must appeal to maximally central cognition: high-level thoughts, theories, and beliefs about objects. Below we review three examples of such arguments, based on (a) the necessity for discrete object representations, (b) the necessity for representations which are not tied to retinal images, and (c) the fact that in some cases perceptual systems seem to violate the very constraints which comprise initial knowledge in infancy.

3.1 The Object Individuation Argument

The first reason for thinking that parts of the infant's object concept cannot be explained in maximally sensory terms is that perceptual systems are thought to be intrinsically *continuous* in nature, and do not map distinct objects to distinct representations.

> Perceptual systems do not package the world into units. The organization of the perceived world into units may be a central task of human systems

> of thought. . . . The parsing of the world into things may point to the essence of thought and to its essential distinction from perception. Perceptual systems bring knowledge of an unbroken surface layout. . . . (Spelke, 1988b, p. 229)

Perception, in other words, does not individuate discrete *objects*, whereas infants "are predisposed to interpret the physical world as composed of discrete, individual entities when perceiving spatial layouts" (Wynn, 1992, p. 750). And again: "[T]he ability to apprehend physical objects appears to be inextricably tied to the ability to reason about the world. Infants appear to understand physical events in terms of a set of principles that guide . . . the organization of the perceived world into units" (Spelke 1988b, p. 198). "*Thought*, in contrast, breaks this continuous layout into units — into objects and events — and finds relations between these units" (Leslie, 1988, p. 201). Since perception doesn't represent discrete objects, and 'thought' does, the correct explanations must appeal to the latter.

3.2 The Occlusion Argument

A second factor which militates against maximally sensory explanations is that perceptual representations are thought to be intrinsically *fleeting* in nature, active only when the corresponding objects in the world are actually visible on the retinae. For example, the sorts of discriminative abilities described above (involving objects traveling behind occluders) must be due to a mechanism which, unlike 'perception', "organizes events in ways that extend beyond the immediately perceivable world in space and time" (Spelke, 1988a, p. 180). "[T]he mechanism appears to carry infants beyond the world of immediate perception, allowing them to make sense of events in which objects are completely hidden" (Spelke, 1988a, p. 172). Contemporary writers still give credence to this argument: Wellman and Gelman (in press) note that these sorts of experiments are "designed to tap *conceptions* about objects, not just object *perception*, in that [they assess] infants' expectations about unseen events — the object's unwitnessed path of movement behind the screen" (our emphases), and Bertenthal (1996) notes in a recent review that "these abstract representations about the motions of objects are accessible to infants as explicit knowledge. . . . The principle evidence for this knowledge derives from occlusion studies in which inferences are required because the entire event is not visible" (p. 450). The responsible mechanisms must therefore "carry infants beyond the world of immediate perception" (Spelke, 1988a, p. 172). When this issue is approached from a dichotomy between perception and cognition, this argument favors the latter: since perceptual representations only exist while their objects are visible, while 'conceptual' representations are free to persist indefinitely, the correct explanations must be maximally central in nature, and appeal to the latter.

(This concern with occlusion, of course, has always been a part of research on the infant's object concept. Piaget, 1954, for instance, held that an important aspect of a 'true' object concept was an ability to represent the locations of objects which were fully hidden from view.)

3.3 The 'Pulfrich Pendulum' Argument

A third sort of argument against perceptual explanations consists in showing that perceptual systems sometimes violate the very constraints they would have to explain. Leslie (1988) proffered this sort of argument against perceptual explanations of the contact-mechanical principle of 'solidity', by which infants seem to have initial knowledge of the principle that objects cannot occupy the same place at the same time, and thus cannot pass through each other (e.g. Baillargeon et al., 1985). To rule out a perceptual explanation of this type of initial knowledge, Leslie identified a type of evidence which would be conclusive:

> Much better would be evidence that input systems are actually quite happy with the idea of one object passing through another. . . . The following kind of evidence is needed: a robust and clearly describable illusion in which one solid rigid object is seen to pass through another solid rigid object; the illusion arises from the visual system's attempt to resolve an incongruity; and it occurs despite the continuous availability of perceptual information that conflicts with the resolving (illusory) percept. (Leslie, 1988, pp. 196 - 197)

It turns out that this type of evidence actually exists, in the form of the *Pulfrich Double Pendulum Illusion* (see Figure 5).

Wilson and Robinson (1986) constructed a display in which two pendulums (sand-filled detergent bottles attached to rigid metal rods) swing back and forth in parallel but in opposite phase, one slightly behind the other. This event is entirely ordinary, unless viewed in dim light (by both eyes), while wearing a neutral density filter over one eye. Such a filter reduces the luminance for one eye, resulting in slightly slower processing. This leads to a percept in which a pendulum's perceived depth varies with its location and direction of motion. A single pendulum is thus seen as swinging in an ellipse, while two pendulums swinging in opposite phase are seen as following intersecting ellipses (Wilson & Robinson, 1986; see Figure 5). Leslie (unpublished) replicated this effect, and subsequently verified that subjects receive "the clear perception of the rigid solid rods passing through each other. Most observers were able to find an angle of view where even the pendulum bottles appear to pass through one another despite their large size and marked surface texture" (Leslie, 1988, p. 199).

The fact that the perceptual systems are 'willing' to construct this sort of percept suggests that they are not the architectural locus of the solidity constraint, which is being blatantly violated. The intended conclusion of this demonstration is that perceptual systems cannot be responsible for infants' initial knowledge of the solidity constraint, since they themselves appear to violate this constraint in constructing the percept of the Pulfrich Pendulum. Again, the implication is that the responsible mechanisms must actually be maximally central.

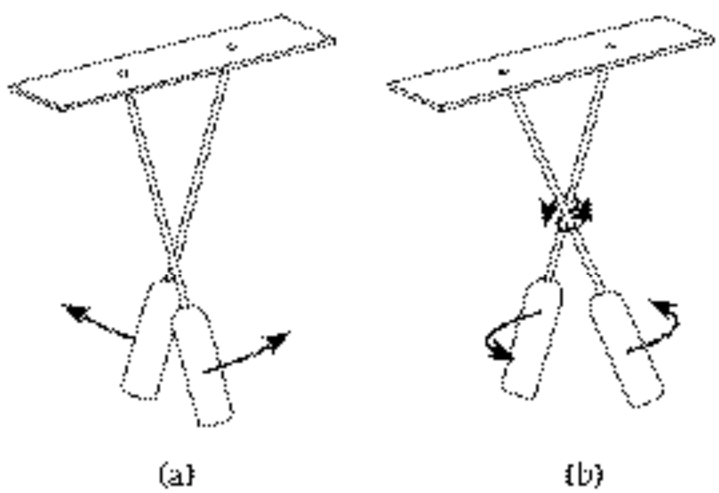

FIG. 5. The 'Pulfrich Pendulum' illusion. (a) What really happens. (b) What is perceived. See text for details. (Adapted from Leslie, 1988)

3.4 The Maximally Central View of the Infant's Object Concept

The impact of these sorts of arguments against perceptual explanations is apparent today in the popularity of explaining these facets of the infant's object concept by appeal to innate *knowledge* and the like. Once more, though, it is precisely the nature of this knowledge which we are discussing. 'Initial knowledge' is not a technical term, and there is nothing inherently objectionable or controversial about suggesting (for example) that "the infant's mechanism for apprehending objects is a mechanism of thought: an initial theory of the physical world" (Spelke, 1988a, p. 181). However, many interpretations of looking-time data are obviously intended to be maximally central in nature. Baillargeon, for example, attributes to Leslie and Spelke the view that "infants are born . . . with substantive beliefs about objects" (1995, p. 184; also 1994, p. 133), and recall Spelke's suggestion that our initial knowledge comprising the object concept is acquired "in ways like those by which we come to know about [an object's] material composition or its market value" (Spelke, 1988b, p. 198). (See also Gopnik and Meltzoff, 1997, ch. 4, for another explicitly maximally central theory of the infant's object concept.)

In sum, several traditional arguments have suggested that perceptual systems cannot be responsible for initial knowledge in infancy, and in the context of the perception / cognition dichotomy this militates in favor of 'cognition' and 'thought'. We are in general agreement with these arguments, and take our task to be an explication of exactly what 'thought' amounts to in this context. We disagree, however, that the responsible mechanisms must therefore be maximally central, and cannot "be distinguished from thought in any sense". In our view, the foundation of spatiotemporally-based initial knowledge lies in neither maximally sensory nor maximally central mechanisms, but rather at the attentional interface between these levels . . .

4. Beyond the Perception/Cognition Dichotomy: Object-Based Mechanisms of Visuospatial Attention

The arguments reviewed in the previous section require that the mechanisms responsible for the infants' discriminative abilities be able to 'parse' the visual world into discrete objects, employing representations which survive occlusion. Maximally central mechanisms certainly meet these requirements. We are struck, however, by the fact that there exist independently-motivated mechanisms of object-based visuospatial attention which also meet these constraints, but which appear to be neither fully 'perceptual' nor fully 'conceptual' in nature. In short, we may have been misled by this artificial dichotomy between 'perception' and 'thought' into thinking that the answer must lie fully at one extreme or the other.

Below, we describe recent object-based conceptions of visuospatial attention, focusing on two theories which will be especially relevant for our purposes: object file theory and the FINST theory of visual indexing.

4.1 The Shift to Object-Based Conceptions of Visuospatial Attention

Attention imposes a limit on our capacity to process visual information, but it is not clear at the outset what the correct *units* are for characterizing this limitation. It was traditionally argued or assumed that attention simply restricts various types of visual processing to certain spatial areas of the visual field — for example in the popular *spotlight* models of visual attention (e.g. Eriksen & Hoffman, 1972; Posner, Snyder, & Davidson, 1980), or the *zoom-lens* metaphor of Eriksen and St. James (1986). It has recently been demonstrated, however, that there must also be an *object-based* component to visual attention, in which attentional limitations are characterized in terms of the number of preattentively-defined discrete *objects* which can be simultaneously processed.

There now exist several demonstrations of object-based effects in visuospatial attention. These include:

- Demonstrations that it is possible to pay attention to distinct objects or 'object schemas' while ignoring other stimuli (comprising different objects) which happened to be spatially superimposed or overlapped (e.g. Neisser & Becklen, 1975; Rock & Gutman, 1981)

- Experimental demonstrations that it is easier to attend to multiple parts of a single object than to multiple parts of two distinct objects, even when the 'parts' in question reside in identical spatial locations, and when the difference is defined only by perceptual set (e.g. Baylis & Driver, 1993; Duncan, 1984)

- Experimental demonstrations that attention automatically spreads more readily from one part of an object to another part of the same object, versus another part of a different object — again, even when the 'parts' in question are spatially identical (e.g. Egly, Driver, & Rafal, 1994)

- Experimental demonstrations that attentional phenomena such as *inhibition of return* and the *negative priming effect* adhere to objects rather than (or in addition to) locations (e.g. Tipper, Brehaut, & Driver, 1990; Tipper, Driver, & Weaver, 1991)

- Corroborating neuropsychological evidence that the phenomenon of *unilateral spatial neglect* operates (at least in some cases) in object-centered rather than scene-centered coordinates, so that patients neglect halves of multiple objects at different locations in the visual field, rather than half of the visual field as a whole (e.g. Behrmann & Tipper, 1994; Driver & Halligan, 1991)

- Additional corroborating neuropsychological evidence from *Balint Syndrome*, in which patients exhibit *simultanagnosia*, the inability to perceive more than one object at a time (for a review, see Rafal, 1997)

For reviews of the recent turn to object-based conceptions of visuospatial attention, see Egeth and Yantis (1997) and Kanwisher and Driver (1992).

Several recent theories have been concerned with how visual objects are individuated, accessed, and used as the basis for memory retrieval. These theories include Kahneman and Treisman's *Object File* theory (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992), Pylyshyn's *FINST* theory of visual indexing (Pylyshyn 1989, 1994), Yantis' *Attentional Priority Tags* (Yantis & Johnson, 1990; Yantis & Jones, 1991), the notion of *Object Tokens* (Chun & Cavanagh, 1997; Kanwisher, 1987), and Wolfe's theory of *Preattentive Object Files* (Wolfe & Bennett, 1997).

To get the flavor of these theories, and how they enjoy the relevant properties which rule out 'perceptual' accounts of the infant's object concept, we will now describe two of these theories in some detail. (These will be central in our theory of *Object Indexing*, presented in Section 5.)

4.2 The FINST Theory of Visual Indexing

Pylyshyn's theory of *visual indexing* (e.g. Pylyshyn, 1989, 1994) complements other theories of object-based attention by postulating a mechanism whereby preattentive object-based individuation, tracking, and access are realized. In order to detect even simple geometrical properties among the elements of a visual scene (e.g. being collinear, or being 'inside'), Pylyshyn argues that the visual system must be able to simultaneously reference — or 'index' — multiple objects. Similarly, although focal attention may scan about until it finds objects, it cannot orient directly to a particular object which has not already been indexed. These considerations suggest a need for multiple loci of 'attention'.

A more concrete demonstration of these requirements is the *multiple-object tracking* (MOT) paradigm. In the standard MOT task (Pylyshyn & Storm 1988; Yantis, 1992), subjects must track a number of independently and unpredictably moving identical items in a field of identical distractors. In the original experiment, Pylyshyn and Storm (1988) introduced the MOT paradigm as a direct test of the visual indexing theory. Subjects in their first experiment viewed a display initially consisting of a field of identical white items. A certain subset of the items were then flashed several times to mark their status as targets. All of the items then began moving independently and unpredictably about the screen, constrained only so that they could not pass too near each other, and could not move off the display. At various times during this motion, one of the items was flashed, and subjects pressed keys to indicate whether the flash had been at the location of a target, a non-target, or neither. See Figure 6 for a schematic representation of this basic MOT task. Since all items were identical during the motion interval, subjects could only succeed by picking out the targets when they were initially flashed, and then tracking them through the motion interval. Subjects were successful (never less than 85.6% accurate) in these experiments when tracking up to five targets in a field of ten identical independently and unpredictably moving items.[7]

All of these considerations suggest the need for multiple loci of attention which can serve to independently 'index' and track a number of salient items. The 'FINST' model of visuospatial attention provides just such a mechanism. Pylyshyn's model is based on *visual indexes* which can be independently assigned to various items in the visual field, and which serve as a means of *access* to those items for the higher-level processes that allocate focal attention. In this regard, they function rather like pointers in a computer data structure: they reference certain items in the visual field (identifying them *as* distinct objects), without themselves revealing any properties of those objects.

Pylyshyn initially called these indexes *FINSTs*, for FINgers of INSTantiation, due to the fact that physical fingers work in an analogous way: "Even if you do not know anything at all about what is located at the places that your fingers are touching, you are still in a position to determine such things as whether the object that finger number 1 is touching is to the left of or above the object that finger number 2 is touching. . . . [T]he access that the finger contact gives makes it inherently possible to track a *particular* token, that is, to keep referring to what is, in virtue of its historical trace, the *same* object" (Pylyshyn, 1989, p. 68).

Visual indexes can be assigned to objects in the visual field regardless of their spatial contiguity (in contrast with spotlight models), with the following restriction: the architecture of the visual system provides only about *four* indexes. Furthermore, the indexes are *sticky*: if an indexed item in the visual field moves, the index moves with it. The visual indexes confer a processing priority to the indexed items, insofar as they allow focal attention to be shifted to indexed (and possibly moving) items without first searching for them by scanning through the intervening space. (Note that the visual indexing theory thus complements — rather than competes with — theories that posit a single locus of focal attention; cf. Pylyshyn & Storm, 1988, p. 180). Attention is typically thought to improve various sorts of low-level visual processing, speeding response times to attended objects or areas (e.g. Downing & Pinker, 1985; Posner et al., 1980). Similarly, visual indexes confer a processing advantage to the indexed items, since they can be immediately accessed by higher-level processes without a serial search. Intriligator (1997, Experiment 2) and Sears and Pylyshyn (in press) explored these issues in the context of multiple-object tracking, demonstrating that this type of processing advantage is target-specific; in particular, it doesn't hold for non-targets — even those located within the convex polygon bounded by the moving targets. Thus, it

---

[7]Pylyshyn and Storm ruled out a class of alternate explanations for this result in which a single spotlight of attention sequentially and repeatedly visits each item in turn: even at the fastest reported scan velocities (around 250 deg/s), a simulated attentional spotlight, augmented with several location-prediction and guessing heuristics, was unable to approach the actual performance of human subjects. See Pylyshyn and Storm (1988) for the details of this simulation.

must be the items *themselves* which are being indexed and tracked in the MOT task, and not the region of space in which they're located.

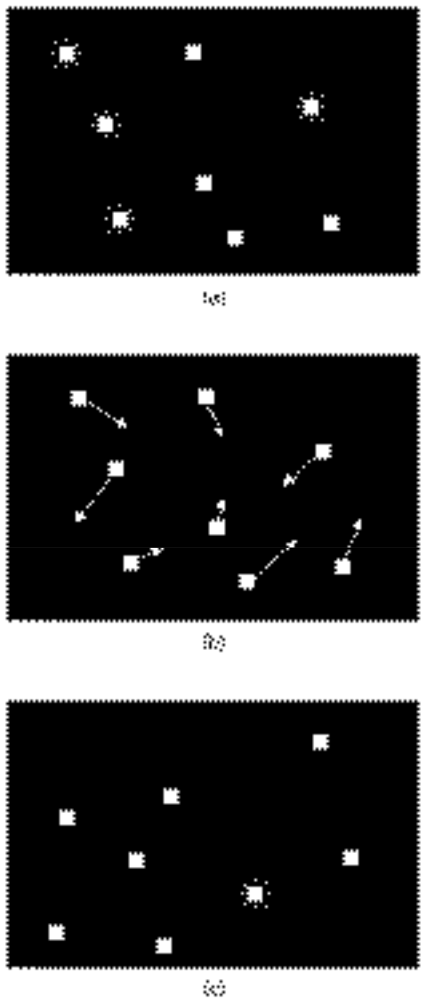

FIG. 6. A schematic depiction of a generic multiple-object tracking task (not to scale). A number of items are presented, and a subset of them are flashed several times to indicate their status as targets. All of the items then begin moving randomly and unpredictably about the screen. At one or more predetermined intervals, the motion stops, and one of the items is flashed again to indicate its status as the probe. Subjects are to decide if the probe item is one of the target items, and respond appropriately.

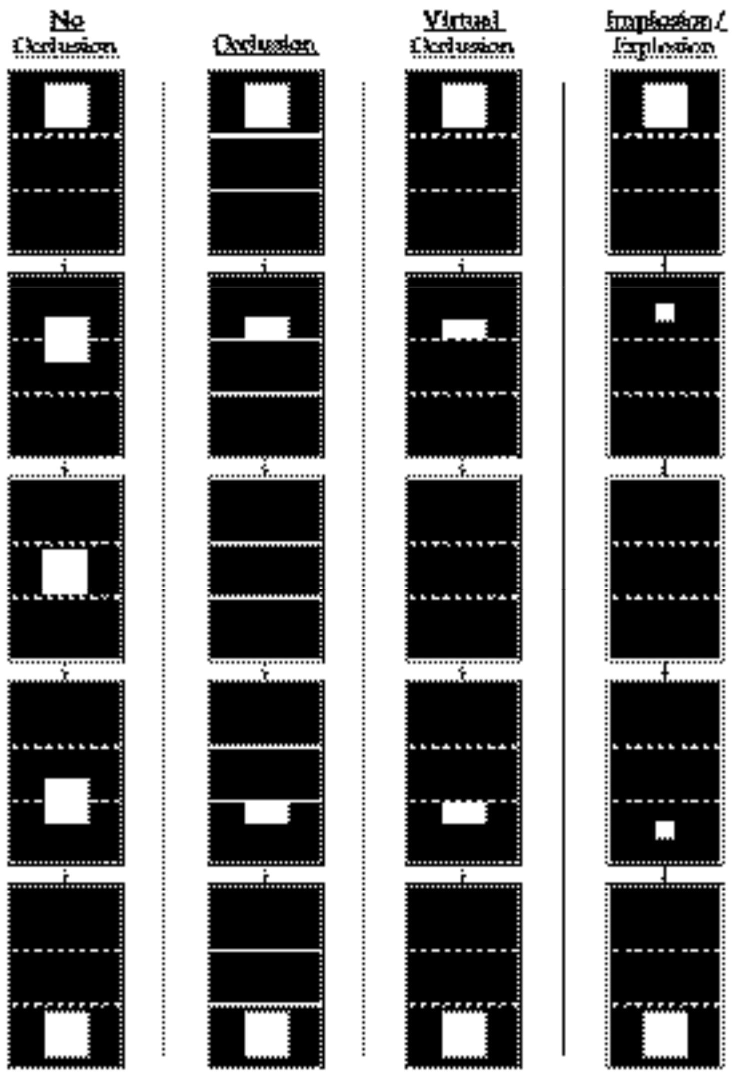

FIG. 7. Some of the different 'occlusion' conditions from Scholl and Pylyshyn (in press). The inherently dynamic nature of the occlusion conditions makes them difficult to represent in a static medium, but here we present some of them as sequences of static 'snapshot' diagrams. In each condition, an item travels downward throughout five sequential frames of motion, interacting with a hypothetical occluder position (not to scale). Solid occluder boundaries represent visible occluders, while dashed occluder boundaries represent invisible occluders (presented to aid comprehension). (Adapted from Scholl & Pylyshyn, in press)

Of particular relevance here is the recent demonstration that visual indexes survive occlusion. Scholl and Pylyshyn (in press) used the standard MOT task but had different conditions where occluders were (visually or functionally) present on the screen. Subjects were able to track items even when the items were briefly (but completely) occluded at various times during their motion, suggesting that occlusion is taken into account when computing enduring perceptual objecthood (see also Section 4.4.2). Unimpaired performance required the presence of accretion and deletion cues along fixed contours at the occluding boundaries. Performance was significantly impaired when items were present on the visual field at the same times and to the same degrees as in the occlusion conditions, but disappeared and reappeared in ways which did not implicate the presence of occluding surfaces (e.g. by imploding and exploding into and out of existence, instead of accreting and deleting along a fixed contour). (See Figure 7 for a schematic depiction of these types of conditions.) This suggests that the visual indexing system is making allowances for occlusion *qua* occlusion, and is not merely robust in the face of any modest interruptions in spatiotemporal continuity.

Several additional experimental paradigms have been used to adduce support for the visual indexing framework, including evidence from *subitizing* (Trick & Pylyshyn, 1993, 1994), *visual search* (Burkell & Pylyshyn, 1997), and the *line-motion illusion* (Schmidt, Fisher, & Pylyshyn, 1998). For concise reviews of this experimental support, see Pylyshyn (1994) and Pylyshyn et al. (1994).

4.3 Object Files

Like Pylyshyn's visual indexing theory, Kahneman and Treisman's *object file* theory attempts to describe the nature of object-based representations of visual attention. We assume, as do Kahneman and Treisman, that visual indexes and object-files are both parts of a single indexing system. Kahneman et al. (1992) suggest that "We might think of [a visual 'FINST' index] as the initial spatiotemporal label that is entered in the object file and that is used to address it. . . . [A] FINST might be the initial phase of a simple object file before any features have been attached to it" (p. 216).

One traditional model of visual experience contends roughly that visual stimuli are identified as objects when their visual projections activate semantic representations in long-term memory. Visual experience, then, consists in shifting patterns of this type of LTM activation. Kahneman et al. (1992) call this the 'display-board model of the mind', and note that it has a number of serious shortcomings. It appears to be the case, for instance, that objects can be perceived and tracked through space even when they remain unidentified. Furthermore, when objects are initially mis-identified, and later correctly recognized, there is still never any doubt that the object involved was the same object. "Two identical red squares in successive fields may be perceived as distinct objects if the spatial/temporal gap between them cannot be bridged, but the transformation of frog into prince is seen as a change in a single visual object" (Kahneman et al., 1992, p. 179). Identification *of* a particular object, in other words, is distinct from identification *as* an object in the first place.

Kahneman and Treisman have argued that an intermediate level of representation is needed to mediate this latter task. In their theoretical framework, this role is played by *object files.* According to their theory, attending to an object in the visual field causes a temporary representation called an object file to be created. This object file stores information about the object's properties (including its color, shape, and current location), and this information is continually updated when the world changes. Object files are allocated and maintained primarily on the basis of spatiotemporal factors, however. Each time an object's spatiotemporal properties change (e.g. in item motion or apparent motion), the new state of the object is compared with the previous state of the object file. If these two states are spatiotemporally similar enough, then the object is seen as continuous, and the object file is updated appropriately. If the two states are sufficiently spatiotemporally dissimilar, however, the previous object file decays, and a new object file is opened to represent the 'new' object.

Kahneman et al. (1992) describe three operations which are involved in managing object files: (a) a *correspondence* operation, which determines for each object whether it is novel, or whether it moved from a previous location; (b) a *reviewing* operation, which retrieves an object's previous characteristics, some of which may no longer be visible; and finally (c) an *impletion* operation which uses both current and reviewed information to construct a phenomenal percept, perhaps of object motion. Kahneman et al. (1992) demonstrated that object files survive real and apparent motion, and Scholl and Pylyshyn (in press) suggest that they also survive occluded motion.

4.4 Object-Based Attention and the Anti-'Perceptual' Arguments

These and other object-based mechanisms of visuospatial attention are not thought to be entirely continuous with 'thought' and 'cognition' in general, but rather to occupy a distinct part of the cognitive architecture which serves as an interface between early visuospatial processing and cognition (Julesz, 1990; Pylyshyn, 1999, in press). They lie, in other words, in neither extreme of the 'perception'/'cognition' dichotomy, but somewhere in the middle. And although these mechanisms have been motivated and developed completely independently from concerns about the infant's object concept, they seem, at a minimum, to enjoy those crucial properties which have sometimes been taken to rule out 'perceptual' explanations, such as individuating distinct objects, and employing representations which survive occlusion.

4.4.1 Attending to the Object Individuation Argument

The first argument that we discussed concerned the requirement that the responsible mechanisms employ discrete representations of individual objects — in contrast to the intrinsically continuous nature of perception. Object-based mechanisms of visuospatial attention meet this constraint. As their very name indicates, these mechanisms are designed to represent discrete objects. Indeed, the object-based attention literature was originally motivated by just this sort of distinction, between continuous spatial representations and discrete object-based representations.

Similarly, both the object file and visual indexing frameworks were explicitly designed for indexing discrete objects, in an attempt to explain exactly how and when discrete objects are represented by mechanisms of visuospatial attention. The existence of such mechanisms casts doubt on the inference from the necessity of discrete representations to the necessity of appealing to maximally central thought to explain the infant's competence. It may be true that 'thought', unlike 'perception', "packages the world into units", but so do mechanisms of object-based visuospatial attention, which reflect parts of cognitive architecture which are not maximally central.

4.4.2 Attending to the Occlusion Argument

The second argument concerned the need for representations which could survive total occlusion — in contrast to the fleeting, retinally-bound nature of perceptual representations. While persistence in the absence of retinal stimulation is perhaps not a necessary feature of object-based representations, it is a feature of both the object file and visual indexing theories. This is most explicit in the case of the visual indexing, as discussed above. Moreover, Scholl and Pylyshyn (in press) argued that their results also had to be interpreted under the object file framework as involving object files persisting through occlusion, and Treisman has confirmed that no upper bound has been experimentally determined concerning how long object files can persist during a cessation of retinal input (personal communication, 1997).

In fact, the persistence of representations through occlusion may be a general theme of attentional processing. There are several reasons to think that allowances for occlusion *should* characterize early visual processing, based on the fact that occlusion permeates visual experience. As Nakayama and his colleagues have noted,

> "[O]cclusion varies greatly, depending on seemingly arbitrary factors — the relative positions of the distant surface, the closer surface, the viewing eye. Yet, various aspects of visual perception remain remarkably unimpaired. Because animals, including ourselves, seem to see so well under such conditions and since this fact of occlusion is always with us, it would seem that many problems associated with occlusion would have

> been solved by visual systems throughout the course of evolution. (Nakayama & Shimojo, 1990, quoted in Nakayama, He, & Shimojo, 1995, p. 62)

These sorts of general considerations have been borne out by Scholl and Pylyshyn's (in press) experiments, and also by several recent related demonstrations. Many of these experiments involve static stimuli in the context of visual search (e.g. Davis & Driver, 1994; Enns & Rensink, 1998; He & Nakayama, 1994), but similar results have been reported in other paradigms involving dynamic stimuli (Shimojo & Nakayama, 1990; Tipper, Brehaut, & Driver, 1990; Yantis, 1995). We will not discuss these studies here for lack of space, but the general lesson of these experiments is that "preattentive processes do more than simply register and group together elementary properties of the two-dimensional image — they are also capable of determining properties of the corresponding three-dimensional scene" (Enns & Rensink, 1991, p. 346). Compare this to Spelke, Vishton et al. (1995) on the relevant occlusion constraint in cognitive development: "[I]nfants' perception of objects depends on analyses of the arrangements and motions of surfaces in the three-dimensional visual layout, not on analyses of the arrangements and motions of the elements in any two-dimensional retinal projection of the layout. . . . [This] suggest[s] that the processes underlying object perception occur relatively late in visual analysis" (p. 168). Spelke and her colleagues are explicitly assuming that properties of the 3D scene are computed only in *late* visual processes, whereas the evidence cited above demonstrates that such computation is performed in early, preattentive processing.

In sum, some preattentive and attentive mechanisms do recognize occlusion, and make allowances for it. In such situations, perceptual objecthood is continuously maintained throughout an item's trajectory via some internal representation — such as a visual index or an object file — even though the object may frequently disappear completely from the visual field.

4.4.3 Attending to the 'Pulfrich Pendulum' Argument

The point of the third argument was that in some cases 'perceptual' systems seem to violate the very constraints that they are being asked to explain. Above, we offered the example of the 'Pulfrich Pendulum', in which the visual system constructs a percept which seems to violate the contact-mechanical 'solidity constraint', about which infants seem to enjoy initial knowledge. The point we wish to stress is that no case has been found in which the attentional system violates the spatiotemporal constraints with which this chapter is concerned, as opposed to contact mechanical constraints (see Section 2.1).

We believe that there may be a fundamental, architecturally real distinction between the spatiotemporal and contact-mechanical types of initial knowledge, and that the former may be explained by appeal to mechanisms analogous to object-based mechanisms of visuospatial attention (see Section 5). According to the logic of the 'Pulfrich Pendulum' argument, we should thus be unable to demonstrate violations of the relevant spatiotemporal principles in the operation of the attentional mechanisms themselves. And indeed, there do not seem to be cases like this in which the attentional systems interpret dynamic scenes in ways which violate basic laws of spatiotemporal continuity, object individuation, and arithmetic. (This is despite the fact that certain contact-mechanical constraints are interchangeable with certain spatiotemporal constraints, for example solidity and continuity; see Spelke et al., 1992). Nor should this seem surprising, since the hypothesized purposes of many attentional mechanisms are exactly analogous to spatiotemporally-based initial knowledge. To foreshadow our discussion in Section 5, consider:

- <u>Spatiotemporal continuity</u>
  Infants seem to have initial knowledge concerning the principle that objects must trace spatiotemporally continuous paths through space, while mechanisms such as visual indexes are designed in the first instance to represent visuospatial objects as they trace spatiotemporal paths through space. Thus when a tracked item in the visual indexing framework disappears from the visual field, the index which was tracking it becomes de-assigned (unless the disappearance is accounted for by an occluding surface, just as in the infancy research).

- <u>Subitizing</u>
  Infants seem to have initial knowledge concerning basic arithmetical operations, so long as the cardinality of the operands is less than about five. This is exactly the limit which has been independently motivated in the visual indexing framework, and indeed one of the primary sources of experimental support for visual indexing comes from studies of the ability of adults to *subitize*, or rapidly determine the cardinality of sets with fewer than 5 items (Trick & Pylyshyn, 1993, 1994).

- <u>The Primacy of Spatiotemporally-Based Object Individuation</u>
  Infants seem to have initial knowledge of how to individuate objects based on spatiotemporal information, which is exactly what mechanisms such as visual indexes and object-files are designed to do — to account for the individuation and tracking of objects, regardless of their (perhaps tenuous) properties. Indeed, the performance of ten-month-olds in Xu and Carey's (1996) experiments exactly mirrors the motivation for mechanisms such as object files.

(We return to these three analogous aspects in more detail below, in Section 5.) While it could be that phenomena analogous to the 'Pulfrich Pendulum' could yet be observed in the case of spatiotemporal constraints, we are not aware of any at present, and we predict that no such violations will be found.

To summarize where we have come: The nature of initial 'spatiotemporal' knowledge of the object concept has sometimes been addressed in terms a dichotomy between perception and cognition, and given this framework there are several arguments which militate in favor of cognition. The resulting explanations of initial knowledge have sometimes attributed to the infants various beliefs and theories which supply the relevant principles about how objects must behave. Such maximally central explanations are not mandated by the traditional arguments, however, since, as we have seen, other mechanisms which are neither maximally central nor maximally sensory find the traditional anti-perceptual arguments entirely congenial. (In the years since the traditional anti-perceptual arguments were first formulated, Spelke has continued to develop her ground-breaking views about the object concept, and seems now to agree, at least, that these two literatures have been developed in similar ways. In recent writings, she keeps the door open to the relevance of the attentional interface, but she does not develop these connections; cf. Spelke, Gutheil et al., 1995, Section 8.5.2.) All of this suggests that it would be interesting to pursue a theory of the object concept which was modeled after such attentional mechanisms rather than after maximally central aspects of objects such as their "market value".[8]

5. Object Indexing Theory

Leslie et al. (1998) introduced the *Object Indexing* theory as an attempt to embody the analogous aspects of the cognitive development and visual attention research programs in a concrete framework. This is not the place in which to mount a detailed defense of the framework, but having emphasized the potential importance of an attentionally-based theory in principle, we hope to sketch a plausible theory in which the insights from the object-based attention literature are put to good use.

---

[8] Another analogous aspect of the infancy experiments and the object-based attentional research lies in the role of *novelty*. Recall that in the looking-time measures, the test phase checks to see if a novel event is interpreted by the infant as a fundamentally new event. The longer looking times, in other words, correspond to perceived *novelty*. This bears a striking resemblance to another recent trend in the study of visuospatial attention, emphasized by Steven Yantis. Yantis has argued that visual attention is automatically captured by the appearance of *new visual objects* (Yantis 1993, 1995). Abrupt onsets, for example, capture attention "not because they are accompanied by a luminance increment, but because they mark the appearance of a new perceptual object" (Yantis, 1993, p. 157). In different contexts (e.g. the reappearance of an object from behind an occluder), the same luminance increment will not be interpreted as a new visual object, and so will not exogenously attract visual attention. Yantis explains this in terms of the underlying responses of the object-file system.

The key notion in the Object Indexing framework is the 'sticky' index. Like Pylyshyn's visual indexes, an object index is an internal representation that is inherently abstract: an index picks out and keeps track of an object by location, without representing any of the object's properties. Once pointing to an object, however, the object can be accessed rapidly, and properties and featural information can be associated with, or 'bound' to an index. An object index functions in the way that a pointer in a computer program might reference a data structure in the computer's memory: it references data without itself revealing any of the data.

To a first approximation, object indexes can do many of the things that pointing fingers can do. Both object indexes and fingers can point to and thus individuate items based on spatiotemporal criteria (e.g. independent motion, the existence of a spatial gap between the items), and can track the continuing identity of an object as it moves about the world. Both object indexes and fingers can also serve to enumerate the number of objects in the world — at least up to the number of available indexes or fingers. At the same time, neither object indexes nor fingers can by themselves reveal an object's color, luminance, composition, or global shape.

As a mechanism of selective attention, the object indexing system is resource-limited, and has only a small number of available indexes. Following Pylyshyn's visual indexing experiments, we predict that there are not more than *four* object indexes, and that this number serves as an effective compromise between focusing resources and being able to compute relations between distinct objects. We hypothesize the following properties of object index assignment (compare Spelke, Gutheil et al., 1995, section 8.2.4):

1. Indexes are assigned primarily *by* location, but not *to* locations. Rather, indexes are assigned to objects in locations.

2. A distinct object can attract only a single index. Multiple indexes cannot be assigned to identical objects.

3. Multiple spatially-separated areas of the visual field may in some cases be assigned a single index if there is no spatiotemporal information (e.g. relative motion) to distinguish them. (Thus, groups with common motion may be assigned a single index.)

4. Once assigned, an index sticks to its object even as the object moves through several different locations. Indexes may follow objects behind occluders, in which case they point to 'somewhere behind the occluder'.

5. When all available indexes are already assigned to objects, a new object can be indexed only by first de-assigning one of the active indexes, flushing its bound features, and reassigning it to the new

object. In this case, the previously indexed object is no longer represented within the object indexing system.

Property information may later drive object indexing, but we assume that this will occur only if spatiotemporal information is absent or ambiguous, and that this process matures only later in development.[9] Featural information, when present, is stored on our model in a *feature map*, which simply registers the *presence* of a feature (cf. Treisman & Gelade, 1980). Crucially, featural information at this stage is registered without any indication of *where* the feature is in the visual field (or what object it is bound to), but only its presence or absence. The appearance of a red cross and a green circle may thus not be distinct on this level from the appearance of a green cross and a red circle, since the same features are present in each case.[10] See Figure 8 for a summary of this model.

The architectural distinction between indexing and feature binding underwrites a functional distinction between spatiotemporally-driven object *individuation* and featurally-driven object *identification*. In processes of visuospatial attention, the former task is accomplished separately and perhaps earlier in online processing than the latter task (cf. Johnston & Pashler, 1990; Pylyshyn, 1989; Quinlan, 1998; Sagi & Julesz, 1985; Tanaka & Shimojo, 1996). In cognitive development, the former task seems to be accomplished separately and perhaps earlier in maturational development (cf. Bower, 1974; Xu & Carey, 1996).

It is tempting to put this all in terms of another traditional distinction, between the processing of 'what' and 'where' — between the processing of which objects are present (regardless of their locations) and where there exist objects (regardless of their identities). Featural and locational information are thought to be processed, largely independently, by distinct anatomical 'streams' in the brain (e.g. Maunsell, 1995; Sagi & Julesz, 1985; Ungerleider & Mishkin, 1984). Talking about 'what' versus 'where' in functional terms is a useful way to characterize our theory of the object concept, but we remain agnostic on the relation of these ideas to the actual relevant neuroanatomical pathways, (a) since it is unclear how these neural circuits map on to the object-based attentional mechanisms, and (b) because the nature of the domains of these circuits has recently become less clear (e.g. Goodale, 1995; see also Bertenthal, 1996).

Why might it make sense to divide up the effort in this way, and to give priority to the processing of spatiotemporal information? One reason may

---

[9] When featural information does begin to play a role in object indexing, it may do so in two importantly-distinct ways: *individuation by feature* and *identification by feature*. These are distinct processes in principle, and also appear to be dissociable experimentally, with feature-driven individuation maturing earlier (for some features) than feature-driven identification (Tremoulet, 1998).

[10] Cf. Wolfe and Bennett (in press) for an intermediate step in this type of process.

simply be *learnability*. Some objects undergo radical featural changes (and may also merely *seem* to do so), while still retaining their identity. Without a reliable

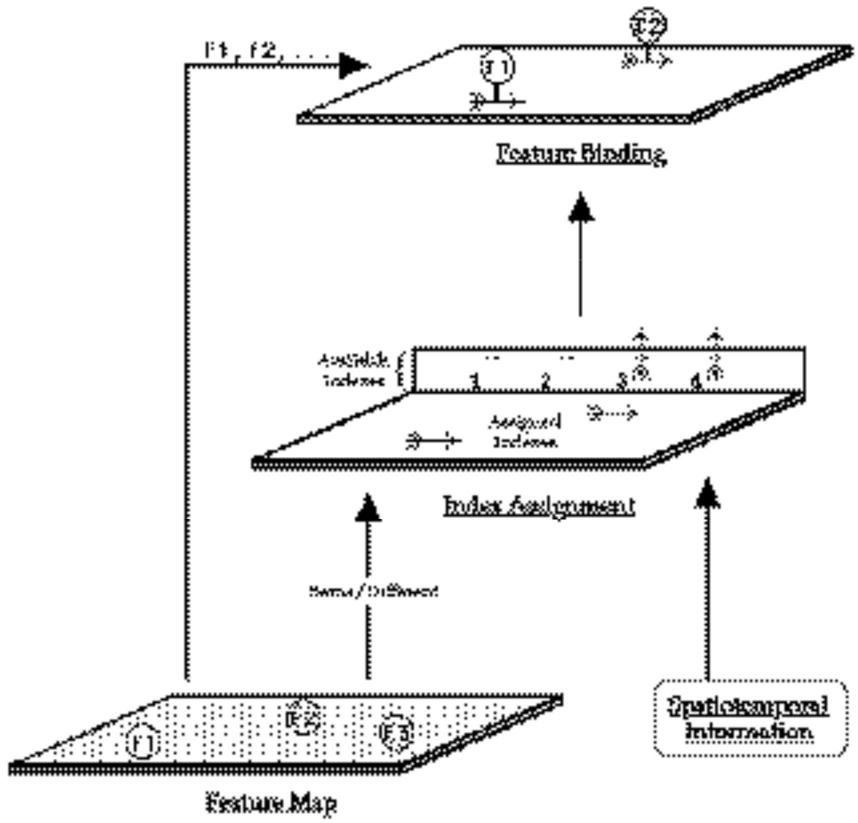

FIG. 8. The mature object indexing system can individuate (i.e. assign indexes) either by location or by feature. The binding of featural information (e.g. F1, F2) to an index (denoted by an arrow) occurs after the index has been assigned, if it occurs at all. Although feature binding is not required for individuation, it is required for identification of objects. We assume that there is only a limited number of available indexes. See text for discussion. (Adapted from Leslie et al., 1998)

guide to which objects engage in such behavior, it may behoove infants to remain agnostic about object identities as determined by featural properties (see also Simon, 1997):

> It might serve the human baby well to use spatiotemporal information to individuate objects, and then slowly learn about the more specific kinds of individuals, and for each which properties change over time and which do not. . . . The infant must learn which properties stay constant for various categories of objects. In order to learn this, however, the infant must have

> some way of individuating objects and tracing them through time. The spatiotemporal criteria for object fulfill this need. (Xu & Carey, 1996, pp. 114, 150)

5.1 What Object Indexing Is Not: Sensations, Principles, Types, and Tokens

Most properties of object indexing are modeled on ideas which were completely independently motivated (e.g. the indexing model, the 'stickiness' of the indexes, the properties of index assignment, the priority of spatiotemporal information, the limit on the number of available indexes, and the feature-binding model). Such a framework differs in several important respects from both maximally sensory and maximally central theories of the object concept.

In the traditional empiricist view, infants first encounter objects in the world only in terms of continuous, fleeting sensations, and only slowly associate these sensations, resulting in bundles of features. Object indexing turns this on its head, and suggests that featural information may often be *ignored* for some purposes early in life, and that the core of the object concept may rather be a kind of mental pointing at a 'this' or a 'that'. In any case, the indexing framework accepts the traditional arguments against maximally sensory views, and assumes that the responsible mechanisms must involve discrete object representations which persist through occlusion. Object indexing is thus opposed to 'perceptual' accounts of the object concept, and we take pains in Section 6.3 to dissociate our framework from contemporary 'perceptual', empiricist, and otherwise deflationary theories (e.g. Bogartz et al., 1997; Munakata, McClelland, Johnson, & Siegler, 1997).

We take object indexing to be continuous with the projects of other researchers such as Carey, Spelke, and Wynn, who have advanced claims of initial knowledge in infancy. The goal of object indexing is to explain what this 'initial knowledge' amounts to, albeit in novel architectural terms. Object indexing does stand opposed, however, to maximally central views of the object concept (see Section 3.4), and tries to steer a course between the maximally sensory and central extremes.

One salient difference between object indexing and maximally central views lies in the fact that any initial knowledge afforded by object indexing would not be in the form of general principles in the way suggested by analogies to scientific theories (Gopnik & Meltzoff, 1997). Several writers have suggested that we think of initial knowledge in infancy in terms of a scientific theory, or an innately driven 'core theory', consisting of general principles (e.g. Gopnik & Meltzoff, 1997). The object indexing mechanisms, in contrast, may be seen as involving 'principles' only in the sense that there are mechanisms of attention that *implement* object principles, without explicitly representing them. The operation of the these mechanisms may indeed conform to certain intelligible physical or numerical principles, but such principles need not be explicitly represented anywhere, any more than "Objects are rigid!" is explicitly stored in the visual system (which treats them as such; e.g. Ullman, 1979). See Pylyshyn (in press) for a general discussion of this distinction.

One way to clarify this is via a distinction between 'types' and 'tokens'.[11] (This terminology is intended only to highlight a crucial distinction involving the *generality* of object representations. In particular, talk of object 'types' in this context has nothing to do with specific *kinds* of objects — e.g. *animals* or *artifacts*.) *Type*-based statements about objects will explicitly quantify over "objects" in general — for example the statements that "Objects trace spatiotemporally continuous paths through space" or that "Moving objects that disappear behind other surfaces with occlusion cues continue to exist behind those surfaces." *Token*-based statements about objects, in contrast, concern only the behaviors and states of affairs of *particular*, individual objects — for example the statement that "*That* object is currently somewhere behind *that* screen."

Object indexing mechanisms account for the infants' abilities without appeal to beliefs or principles about object *types*, but only to reactions to specific object *tokens*. In maximally central accounts, the initial knowledge is thought to take the form of principles concerning object types in general. Infants, for example, know that 'Objects must trace spatiotemporally continuous paths through space' — an explicit statement about objects in general. The object indexing mechanisms, in contrast, may be such that they eventually result in beliefs about individual object tokens (i.e. *that* object must be behind the screen; *that* object must have disappeared; *this* object must be different from that object which I saw a moment ago) without appealing to object types in general. In other words, the relevant conclusions about particular objects fall out of the design of the system, and may conform to the relevant principles without ever explicitly representing them (cf. Kellman, 1988).

Could object indexing account for the results of the infancy experiments which fueled the attributions of initial spatiotemporally-based initial knowledge? In the following sections we return briefly to the three experimental demonstrations of initial knowledge originally discussed in Section 2, and we interpret them in the context of object indexing.

5.2 Indexing and Spatiotemporal Continuity

In the first example, Spelke and her colleagues (Spelke & Kestenbaum, 1986; Spelke, Kestenbaum et al., 1995) demonstrated that infants had initial knowledge of the principle that '*Objects must trace spatiotemporally continuous paths through space*', by examining their looking times while they observed objects moving behind occluders (see Section 2.3 and Figure 1).

---

[11] We are extremely grateful to Susan Carey for helping us to see object indexing in these terms.

The results of these experiments can readily be explained within the object indexing framework. The single item in the first moments of the familiarization displays (Figure 1a and 1b) attracts a single index, which stays attached as it moves along behind occluders (analogous to visual indexes tracking objects behind occluders; Scholl & Pylyshyn, in press). In the 'continuous' condition (Figure 1a), occlusion cues signal the indexing system that the object is behind an occluder. When the item reappears, the single object index reacquires and continues to track it. The event is thus apprehended using only a single index, and infants therefore expect only a single object. The one-object probe (Figure 1c) also employs only a single object index, and so has little novelty. With the two-object probe (Figure 1d), however, increased attention is allocated in the form of a new index, and infants look longer.

In the 'discontinuous' condition (Figure 1b), the first index does not reacquire the object at the screen's occluding boundary, and cannot jump the gap. When an object then appears from the far side of the other occluder, a new index must be assigned. This event requires two object indexes, and is thus apprehended as involving two distinct objects. The 2-object test event (Figure 1d) will then require the same number of indexes, and has little novelty. In the one-object test event (Figure 1c), however, increased attention is allocated to search for the 'missing' object corresponding to the original index (which is still pointing 'somewhere'), and infants look longer.

Spelke concludes from the continuity experiments that "Infants appear to apprehend the identity of objects by analyzing the apparent continuity or discontinuity of paths of motion, in accord with the principle that objects move on spatio-temporally continuous paths" (Spelke, 1988a, p. 179). This is entirely true in the object indexing framework, but the "in accord with" part becomes crucial: maintenance of spatiotemporal continuity is part of the modus operandi of the indexing system (operating over object *tokens*), rather than an explicitly represented principle (about objects as *types*).

5.3 Indexing and Arithmetic

The second example of spatiotemporally-based initial knowledge we discussed above was Wynn's (1992) demonstration that 5-month-old infants can "compute the numerical results of . . . arithmetical operations" (p. 750) (see Section 2.4 and Figure 2). Object indexing theory can explain these results by appeal to the fact that the indexing system can track the numerosity of objects in a scene simply by assigning indexes — rather than by counting to determine a cardinal value.

Infants' looking times in Wynn's '1+1 = 1 or 2' condition are explained as follows. The original item draws an object index, which continues to point to 'somewhere behind the screen' when the object is occluded. The new item then appears in a location remote from the screen, attracting a new index. When it too disappears behind the screen, the system has two indexes both pointing to 'somewhere behind the screen'. These two active indexes translate into an attentional expectation for 2 objects when the screen is lowered. If only a single object remains on the stage, then one active index is left object-less, and increased attention is allocated in a 'search' for the missing object. This increased attentional allocation yields longer looking times.

Similarly in the '2-1 = 1 or 2' case: two indexes are assigned initially to the visible and distinct pair of objects, and then point to 'somewhere behind the screen' when they are occluded. When the single object is removed, one of the active indexes reacquires it, and tracks it off the stage. The lone remaining active index reacquires an object as soon as the screen is lowered, but a new index must be assigned in the anomalous case because there is an extra (i.e. non-indexed) object present behind the occluder. Attention is allocated to effect this index assignment, which yields longer looking. The same type of explanation holds in the '1+1 = 2 or 3' case. Wynn is correct that the 'precise result' is being computed, but this may be effected implicitly by index assignment, and not explicitly by representing cardinal values. Again, the patterns of looking times in these experiments may simply reflect the modus operandi of the object indexing system, rather than knowledge of general, explicitly represented arithmetical principles.

Object indexing also provides a straightforward account of the upper limit on infants' numerical abilities. On the account just sketched, the infant will only be able to track precise numerosities if they remain in the range of available indexes. We assume — in line with the FINST and object file theories — that there exist no more than about four object indexes. That there should be *some* limit follows from the fact that indexing is a resource-limited mechanism of selective attention. The *particular* limit of about four derives empirically from studies of visual indexing and its relation to subitizing (Pylyshyn, 1994; Trick & Pylyshyn, 1993, 1994).

Finally, object indexing explains the otherwise-nonintuitive result of Simon et al. (1995). Simon et al. (1995) found that five-month-old infants in Wynn's paradigm have expectations regarding the numerosity of occluded objects, but did not seem to have any expectations about the properties or identities of those objects. Again, this is to be expected, if the indexes are assigned and accessed by location, and without feature binding. The assigned indexes are '*feature blind*', and serve only to *individuate* the objects. The different properties of the objects in Simon et al.'s (1995) experiment are not 'visible' to the infant's object indexing system.

In sum, using the object indexing framework to account for the 'arithmetic' results seems quite natural. Indexing theory can account for the numerosity tracking, the apparent upper bound on infants' numerical competence, and the lack of anomalous-property effects, all without appeal to explicitly represented arithmetical principles or cardinal values.

5.4 Indexing and Object Individuation

The third example discussed above was Xu and Carey's (1996) demonstration that ten-month-old infants will infer the existence of two occluded objects on the basis of spatiotemporal information (i.e. seeing both objects simultaneously in different locations), but not on the basis of property / kind information (i.e. seeing a duck and a ball sequentially), whereas 12-month-old infants will infer objecthood on the basis of both sorts of information (see Section 2.5 and Figures 3 and 4). In the object indexing framework, these results exemplify the development of feature-binding.

The ten-month-old's performance can be explained by feature-blind object indexing. In the spatial condition, the infant sees two objects in different locations simultaneously at the beginning of the familiarization phase, which results in the assignment of two object indexes. These indexes continue to point to 'somewhere behind the screen', occasionally re-acquiring one of the objects as it returns to view momentarily. These two active indexes translate into an attentional expectation for two objects, so that when only a single test object is revealed behind the screen, increased attention is allocated in a search for the missing object. Again, this increased attentional allocation engenders longer looking times. In the property condition, a single index is initially assigned when the single object — say, a duck — first emerges from behind the occluder. When the duck returns behind the screen, the index tracks it, and ends up pointing to 'somewhere behind the screen'. When an object — say, a ball — then appears from behind the other side of the screen, the 'feature-blind' index has no information contradicting the assumption that this is the originally indexed object, so the object reacquires the old index, rather than attracting a new index. At the end of the familiarization period, the indexing system thus has only a single index active, and doesn't track the existence of the second object, despite its distinct property / kind information.

By twelve months, however, featural differences can now drive index assignment in this type of situation, where spatiotemporal information is ambiguous (see Figure 8). The presence of novel features on the feature map now indicates to the system that the current object is distinct from the earlier object. Perhaps the development from feature-blind to feature-driven indexing reflects increased integration of 'what' and 'where' systems in the brain (Leslie et al., 1998).

Recent studies have suggested that feature binding may still be fragile at twelve months (Tremoulet, 1998). These results indicate that there is also an important distinction between *individuation* by feature (which appears to develop earlier) and *identification* by feature (which appears to develop later). For details of these experiments, see Tremoulet (1998); for discussion, see Leslie et al. (1998).

Again, it should not be surprising that the object indexing framework can address these results, since it is modeled on mechanisms from the visuospatial attention literature which are specifically posited to do just the sorts of things that infants seem to be doing in these experiments. In developing the object file framework, for instance, Treisman and her colleagues were motivated by the necessity (a) to individuate objects solely on the basis of spatiotemporal information when there is no featural information available, and (b) to maintain a continuing representation of objecthood on the basis of spatiotemporal factors, even when featural information is in flux. Thus, again: "Two identical red squares in successive fields may be perceived as distinct objects if the spatial/temporal gap between them cannot be bridged, but the transformation of frog into prince is seen as a change in a single visual object" (Kahneman et al., 1992, p. 179). Ten-month-old infants in Xu and Carey's (1996) experiment responded precisely in accord with these motivations: (a) they cognized two objects when they saw the objects appear simultaneously, while (b) they did not cognize the existence of two objects when two featurally-dissimilar objects were presented sequentially.

In the preceding pages, we have attempted to motivate the idea that the object indexing framework provides a useful way to think about some spatiotemporal aspects of the infant's object concept. In the following section, we highlight some of the challenges that the theory faces.

5.5 Challenges for Object Indexing

Work on visuospatial attention and cognitive development has uncovered surprisingly parallel phenomena concerning the relationship between the spatiotemporal individuation and featural identification of objects. This suggests that similar mechanisms may be at work in both cases. At this stage, however, we present the object indexing model primarily as a framework to guide research in both fields, and to emphasize the theoretical importance of several open questions. We have left it an open question as to the precise relationship between object indexes and the analogous entities from visuospatial attention such as visual indexes and object files (see Leslie et al., 1998). While it may be that these are merely analogous mechanisms and ideas, it is also possible that the FINST and object file frameworks are continuous with object indexing (Scholl, 1997). At least in this strong form, the proposal faces several challenges, which we identify below.

- Reconciling Time Scales
  The crucial events in looking-time experiments with infants typically take on the order of tens of seconds. The experiments of Wynn (1992) and Xu and Carey (1996), for instance, typically involve objects which are occluded for several seconds at a time. The time-scales at work in the attentional experiments, in contrast, are often on the order of tens or hundreds of milliseconds (e.g. the latency between spatiotemporal individuation and featural identification, or the 'reviewing' process for object files). Further work is necessary to determine if these time-scales can be reconciled. For instance, it would be worthwhile to determine exactly how long an object file stays 'open' when its object goes behind a screen and doesn't emerge. Object files are thought to remain active for "at least 600 - 700 ms, and perhaps much longer" (Kahneman et al., 1992, p. 208), but an upper bound on the persistence of object files has yet to be determined (Treisman, personal communication, 1997).

- Cognitive Penetrability
  Since object indexing is thought to be driven by parts of cognitive architecture which are not maximally central, we might predict that only limited kinds of information influence indexing. We might predict, in other words, that at least in some cases object indexing will be *cognitively impenetrable* (Pylyshyn, 1980, 1984). Some contact-mechanical judgments about objects' behavior do seem to be cognitively penetrable. In Baillargeon's 'drawbridge' experiments (Baillargeon et al., 1985), for instance, infants look longer at a drawbridge which impossibly rotates 'through' a solid object, but this pattern of looking does not occur if infants are given evidence that the object is compressible and not rigid (Bailllargeon, 1987). Do such cases exist for the spatiotemporal aspects of the infant's object concept? We are not aware of any. Indeed, it may be that the spatiotemporal-individuation vs. property-based-identification distinction marks an architectural distinction in terms of '*top-down influences*', including cognitive penetrability (e.g. Heller, 1997; Scholl, 1999).

- Cross-Modality
  An interesting challenge to the object indexing framework comes from cross-modal aspects of the object concept (see Spelke, 1988b). Some contact-mechanical aspects of the object concept, at least, do seem to be cross-modally sensitive (e.g. Streri & Spelke, 1988; Streri, Spelke, & Rameix, 1993), and the same may hold for spatiotemporal aspects (cf. Starkey et al, 1990; Wynn, 1996). In

general: "Objects do not appear to be apprehended by separate visual and haptic mechanisms but by a single mechanism that operates on representations arising either through vision or through touch" (Spelke, 1988a, p. 175). The apparent conflict lies in the fact that mechanisms of object-based attention have primarily been proposed to function over only a *visual* domain, and to be somewhat distinct from other modalities. Of course, nothing precludes similar mechanisms existing in other modalities: it has often been noted that visual and auditory processing, for instance, share several surprisingly similar mechanisms (e.g. Julesz, 1980). In addition, although the visual aspects of Pylyshyn's visual indexing framework that have received the most experimental attention, visual indexes are primarily intended to be part of a visuo-*spatial* system, and Pylyshyn's model explicitly includes a proprioceptive component (Pylyshyn, 1989). It is possible that object indexing operates in a single cross-modal space, but this is not a necessary aspect of our model.

- Other Spatiotemporal Explananda?
  We have aimed the object indexing framework only at those non-contact-mechanical, 'spatiotemporal' aspects of the infant's object concept, and have discussed how it might address several relevant experiments. Are there nevertheless other 'spatiotemporal' aspects of the object concept which resist explanation via object indexing? Certainly there are many other sorts of spatiotemporally-based initial knowledge which we have not examined in this chapter, but at first blush many of them seem like excellent candidates for an explanation in terms of object indexing. One such example is Kellman and Spelke's demonstrations that infants will use common motion to infer the *unity* of dynamic partially-occluded objects, but at the same time (and unlike adults) they will not use static featural information to infer the existence of two separate objects (Kellman & Spelke, 1983; Kellman, Spelke, & Short, 1986; see also Johnson & Aslin, 1995, 1996). Another example is inferring object unity from successive dynamic displays (van de Walle & Spelke, 1996). Are there any examples which do not seem to fall within the domain of object indexing — perhaps involving sensitivity to objects' *heights* (e.g. Baillargeon & Graber, 1987; Baillargeon & DeVos, 1991)? (Are properties like 'height' spatiotemporal or featural in nature?)

- Language and the Object Concept
  One final potential challenge to the object indexing framework may be a link to *language learning*. Xu and Carey (1996, Experiments 4 and 5) reported some provocative preliminary evidence suggesting that infants will individuate objects on the basis of property / kind information only when they know the words which name those objects. Of course, as Xu and Carey note, one cannot infer any causal relation from this pattern. Nevertheless: "The correspondence in time of the two developments — the ability to use the differences between a ball and a bottle to individuate objects and the ability to comprehend nouns such as 'ball' and 'bottle' — raises the question of the relation between the two achievements" (Xu & Carey, 1996, p. 145). Object indexing, as presented, contains no explicit (much less necessary) link to language-learning, and thus could only account for this type of result by appeal to ad-hoc additions to the theory. On the other hand, the same might be said for many other explanations of the object concept, which has traditionally been thought to be functionally divorced from language acquisition. Spelke, for instance, has suggested in the past that "language plays no important role in the spontaneous elaboration of physical knowledge" (Spelke, 1988a, p. 181). It may be that noticing an object kind is a prelude to learning a verbal label for it. A mature object indexing system may in turn be required before a given object can be noticed.

5.6 Summary

Despite — and because of — these challenges, we believe that object indexing provides a refreshing way to think about spatiotemporal aspects of the infant's object concept. We have tried with the object indexing framework to bring together two literatures which have previously been developed independently. The surprising 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, and we have tried to flesh out this intuition in our framework. While our theory is intended at present simply as a framework to guide research into these analogies, we have tried to present a plausible story of how such a theory could actually begin to address some of the experiments which have been used to characterize the infant's object concept.

It seems to be the case, as Spelke suggests, that we come into our initial knowledge of the object concept in ways quite different from how we come to know about an object's properties (e.g. its color and form), but it may also be the case that we acquire initial knowledge about objects in ways quite different from "those by which we come to know about its material composition or its market value" (Spelke, 1988b, p. 198). Our earliest knowledge of objecthood may involve not undifferentiated arrays of sensations (as in maximally sensory views), and not explicitly represented principles about object types (as in maximally central views) but rather a kind of inherently abstract mental 'pointing' at a 'this' or a 'that'.

In any case, the object indexing framework is an example of a theory which attempts to escape the traditional bounds of the perception / cognition dichotomy, to a more motivated middle ground, at the attentional interface.

6. Related Theories

In this section we briefly relate both the perception / cognition dichotomy and the object indexing framework to several other recent related accounts of the infant's object concept.

6.1 Simon's 'Non-Numerical' Account of Numerical Competence

Simon (1997) has offered an account of Wynn's 'arithmetic' experiments (see Section 2.4) which in some respects the object indexing framework finds congenial. Simon argues that the patterns of looking times in Wynn's experiments do not require any specifically numerical abilities, but only domain-general abilities which are being co-opted for a purpose for which they were not specifically designed. These abilities include (1) "the ability to remember and compare items from a previously viewed collection", (2) the ability "to make discriminations between collections of up to four objects", (3) "the ability to form representations that generalize over some or all of the perceptual details of the actual items involved", and (4) the ability to notice the disappearance of an occluded object (Simon, 1997, p. 361). Simon argues that these sorts of general abilities are sufficient to account for Wynn's results, and that in particular no additional ability to compute ordinal relationships is needed. Simon's strategy is to catalog this minimal set of necessary abilities, and then to cite evidence that each ability exists. (Simon, in press, also presents a straightforward computational model of this pattern of abilities, and uses the model to emulate the results of the infants in Simon et al, 1995.)

Like object indexing, Simon's account appears to be neither maximally central nor maximally sensory in nature. There are, however, a number of salient differences between our approaches. First, Simon's theory is intended only to account for the results concerning initial knowledge of number and arithmetic, whereas object indexing addresses all spatiotemporally-based initial knowledge. Even within the domain of number, however, our approaches differ on several counts. While Simon does mention mechanisms like FINSTs and object files, his theory is primarily based on an analysis of the *abilities* required to give rise to the relevant patterns of looking-times, without proposing specific mechanisms. Object indexing, in contrast, is focused directly on the responsible *mechanisms* of cognitive architecture, and thus draws together all of these otherwise-unrelated abilities into a single coherent explanation. For example, Simon appeals to the abilities to subitize and to generalize over perceptual features, whereas we appeal to a specific architectural system which explains these abilities (viz. the visual indexing system; see Section 4.2). Simon has clearly recognized the analogies between the visual attention and cognitive development literatures, however, and it may be possible to read our theory as an architectural explanation of the abilities present in Simon's theory. Again, however, the scope of our indexing theory is greater and is not restricted to the domain of number and arithmetic — only one of the types of initial knowledge addressed by object indexing.

6.2 Sortal Concepts and the Object-First Hypothesis

Xu and Carey (1996) have offered a theory of the infant's object concept based on the notion of *sortal concepts* (e.g. Hirsch, 1982; Wiggins, 1980; Xu, 1997). A sortal, to a first approximation, is a concept that provides criteria for determining object individuation and identification. As an example, Xu and Carey (1996) ask us to consider the question 'How many are there in a deck of cards?'. This question has no well-defined answer, since it doesn't specify what is to be individuated and counted (cards? suits? particles of matter?). As this example suggests, sortal concepts are typically lexicalized as count nouns, at least in those languages which employ a count-noun / mass-noun distinction..

Xu and Carey (1996) explain their results (see section 2.5) by appeal to the notion of sortal concepts, along with the '*object-first*' *hypothesis*. Xu and Carey argue that infants employ the sortal concept BOUNDED PHYSICAL OBJECT before they represent any other more specific sortals.[12] This explains why ten-month-olds only expect two objects behind the final screen when given appropriate spatiotemporal information — because that is the only type of information embodied in their only sortal concept. By twelve months, in contrast, other sortal concepts (e.g. BOTTLE, TRUCK) have begun to develop (along with *language*; see Section 5.5), so that the older infant is able to use these additional sortals to succeed in property / kind conditions, where the spatiotemporal information is ambiguous.

We're intrigued by this proposal because BOUNDED PHYSICAL OBJECT is just the sort of necessarily-abstract representation afforded by object

---

[12] Xu and Carey (1996) do not claim that the BOUNDED PHYSICAL OBJECT sortal is lexicalized for the infant, or indeed in the language at all, but they do claim that it is psychologically real for the infant (Xu & Carey, 1996, fn. 1). For critical discussion of the notion of a PHYSICAL OBJECT sortal, see Xu (1997) and commentaries by Ayers (1997), Hirsch (1997), and Wiggins (1997).

indexing. However, we are not yet sure that this idea is best explained by appeal to sortal concepts. One way to characterize our preference here is in terms of *falsifiability*. Consider what the two theories could have said if the relevant experiments had come out the other way for the 10-month-olds — success in property condition (Figure 4), but failure in spatiotemporal condition (Figure 3). The sortal framework could have just as easily accounted for this pattern of results — Xu and Carey could simply posit an '*object-last' hypothesis* instead of the object-first hypothesis. Object indexing, in contrast, motivated independently by object-based mechanisms of visuospatial attention, has no such recourse. In other words, object indexing *must* predict these Xu and Carey (1996) results, whereas the sortals framework can easily be accommodated to fit nearly any pattern of results. This is an explanatory advantage of the object indexing account.

6.3 Contemporary Empiricist Accounts of the Object Concept

There is always a temptation to group all non-maximally-central views of the object concept together as deflationary tendencies, toward the maximally sensory extreme. However, we also reject the other extreme, and have proposed that the appropriate explanation lies at the attentional interface between perception and cognition. In an effort to clarify the distinction between empiricist sensory views and the attentional/indexing view, we briefly discuss two examples of other recent non-maximally-central theories.

Bogartz et al. (1997) explicitly reject maximally central explanations of the object concept, and argue that the relevant patterns of infant looking times can be explained by appeal to 'perceptual processing', adopting the other extreme: "We assume that young infants cannot reason, draw inferences, or have beliefs" (p. 411). Bogartz and his colleagues replicated a study by Baillargeon and Graber (1987) which they suggest had previously been interpreted in maximally central terms. Baillargeon and Graber's study, which we have not discussed, involved a sensitivity to the *height* of an occluded object. Infants were habituated to both a 'short' and a 'tall' toy rabbit passing behind a tall solid screen, and were then tested with a screen which had a high 'window' cut out of its center. Infants dishabituated to a 'tall' rabbit which did not appear through this high 'window' while passing behind the screen, but did not dishabituate to a 'short' rabbit which did not appear in the same high 'window' while passing behind the same screen. These results were interpreted in terms of initial knowledge of spatiotemporal continuity and how it interacts with the *heights* of objects. (See also Baillargeon & DeVos, 1991, and Section 5.5 of this chapter.) Bogartz et al. (1997) replicated this experiment using a more complex experimental design and statistical analysis (involving regression rather than the analysis of variance), and concluded: "The results show unambiguously that the impossibility of the rabbit not showing up in the window played no role in the looking times of these infants" (p. 418). They take this demonstration to support what they call their 'perceptual processing' view of the infant's object concept. (See Bogartz & Shinskey, 1998, 1999, for recent applications of this view to other experiments.)

Bogartz and his colleagues enthusiastically adopt the dichotomy between 'perception' and 'cognition', noting the potential 'over-interpretation' involved in maximally central, 'cognitive' interpretations, and opposing this with a 'perceptual' story. While we agree with their careful treatment of maximally central views, we find 'perceptual' approaches to be inadequate as well, for all the reasons discussed above in Section 3.

We find it hard to discern what exactly Bogartz et al. mean by 'perception' and how the distinction between 'perception' and 'cognition' is to be drawn. All they tell us is that, "Perceptual processing consists of analysis of the immediate representations of events, the construction in associative memory of transformations of these immediate representations, abstraction of their forms, and the comparison of immediate perceptual representations to the representations stored in memory" (p. 411). They then offer the '*modifiable videotape*' metaphor: "The young infant is primarily engaged in making 'videotapes' of events in the world and storing these tapes for access and updating. When new events are perceived, they are compared with the closest tape in the library. If there is a match, nothing new is entered into the library. If there is a discrepancy, either a new tape is created or the old tape is modified to include the new representation as a permissible variation" (p. 411 fn). Without a serious account of the nature of these representations, the proposal is largely empty. For example, in order to avoid circularity, Bogartz et al. need to characterize 'representations of events' that are 'immediate' independently of characterizing 'perceptual processing'. In any case, how could perceptual processes possibly do *all* these things? Unfortunately, aside from a single allusion to visual mental imagery, Bogartz et al. do not provide a single reference to work on perception or vision supporting the claim that perceptual systems have these abilities. It remains to be seen whether their notion of 'perceptual processing' can be made concrete enough to be useful or falsifiable and whether their operations of 'transformation', 'abstraction', and 'comparison' can really be distinguished from 'cognition'. In contrast to these proposals, object indexing theory makes a set of specific empirical claims about how spatiotemporally-based initial knowledge is embodied in infants and adults at the attentional interface between perception and cognition.[13] (These comments, however, do not impugn the novel methodology employed by Bogartz et al., 1997, which we view as a tremendous improvement over traditional methods.)

---

[13] In a similar vein, Melkman and Rabinovitch (1998) suggest that the results of Spelke et al. (1995) can be explained by appeal only to "sensory or perceptual understanding" (p. 258). However, they do not say what they mean by this, and, like Bogartz et al. (1997), they do not provide a single reference to the vision or perception literature to substantiate their dubious claims about the abilities of 'vision' and 'perception'.

Munakata et al. (1997) also reject maximally central explanations of the object concept, and favor instead an approach based on parallel distributed processing. Munakata et al. focus on the question of why certain types of initial knowledge manifest themselves in some tasks but not in others. In particular, they focus on the fact that looking-time measures seem to reveal accurate representations of the locations of occluded objects, while reaching measures seem to indicate that infants do not possess such representations (e.g. the 'A-not-B error'; Piaget, 1954). They suggest that a maximally central view of the object concept (what they call a 'principle-based approach') is not necessary to explain this pattern of results, since they can be adequately modeled in a connectionist network. Munakata and her colleagues have attempted to model infants' looking and reaching behaviors in a computer simulation by employing two sets of output units. To account for the delay between looking and reaching competence, they simply delayed training on one set of outputs, and used a reduced learning rate. They call this set of outputs 'reaching', and thus model the developmental sequence. The relevant 'knowledge' isn't simply present or absent, but rather exists to different *degrees* in different systems: "the ability to represent occluded objects depends on the connections among relevant neurons and . . . the ability is acquired through a process of strengthening these connections" (p. 689). The idea is basically that a 'weak' internal representation in the network may suffice to guide looking, but may not be strong enough to drive reaching.

These are interesting ideas, though we think the connectionist modeling plays little useful role in the explanation. Suppose the infant had been organized such that initial knowledge was revealed by reaching behaviors *before* being revealed by looking times. This would imply a very different architectural arrangement in the infant, but the Munakata et al. model could accommodate this simply by changing the *label* on the delayed set of outputs from 'reaching' to 'looking'!

Object indexing theory does not yet address the issue of why reaching behaviors show delayed initial knowledge relative to looking times. Bertenthal (1996) has argued for the relevance to infancy of a distinction between 'perception for action' and 'perception for cognition', a distinction motivated independently by findings with adults (e.g. Goodale, 1995). One possibility is that these distinct systems employ different indexing mechanisms. In any case, both object indexing and Munakata's approach agree that the infant's object concept is not maximally central.

7. Concluding Thoughts

In this chapter we have focused on the mechanisms of cognitive architecture which underwrite and explain parts of the infant's object concept. Our project is continuous with those of other researchers such as Carey, Spelke, and Wynn, who have demonstrated a wide array of initial knowledge in infancy. The interesting question is what this initial knowledge amounts to in terms of the underlying cognitive architecture. Traditionally, these issues have been viewed in terms of a sharp dichotomy between 'perception' and 'cognition'. Some theorists have rejected explanations involving 'initial knowledge', and have attempted to explain the discriminative abilities of infants solely in terms of low-level perceptual systems. Other researchers have argued that 'perception' lacks several crucial properties (e.g. object-directedness, and the ability to survive occlusion), and that the responsible mechanisms must therefore be mechanisms of maximally central thought.

We reject these extreme approaches, and instead have explored the idea that the origins of the object concept lie at the attentional interface between perception and cognition. In an attempt to further motivate this type of explanation beyond mere possibility, we have offered our *Object Indexing* framework as a potential incarnation of this type of explanation, and have discussed several other sorts of competing theories. In so doing, we have highlighted the facts that the range of possible architectural interpretations of this exciting research is wider than is sometimes thought, and that there are intriguing parallels between research on the infant's object concept, and research on the nature of object-based visuospatial attention in adults. We welcome the vigorous debate that is clearly beginning about the underlying cognitive mechanisms, since for us this is the point of the whole endeavor.

Rutgers Center for Cognitive Science
Rutgers University

References

Antell, S., & Keating, D. (1983). Perception of numerical invariance in neonates. *Child Development, 54,* 695 - 701.

Ayers, M. (1997). Is *physical object* a sortal concept? A reply to Xu. *Mind & Language,* 393 - 405.

Baillargeon, R. (1987). Young infants' reasoning about the physical and spatial properties of a hidden object. *Cognitive Development,* 2, 179 - 200.

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science, 3,* 133 - 140.

Baillargeon, R. (1995). Physical reasoning in infancy. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 181 - 204). Cambridge, MA: MIT Press.

Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. *Child Development, 62,* 1227 - 1246.

Baillargeon, R., & Graber, M. (1987). Where is the rabbit? 5.5-month-old infants' representation of the height of a hidden object. *Cognitive Development, 2,* 375 - 392.

Baillargeon, R., Spelke, E., & Wasserman, S. (1985). Object permanence in 5-month-old infants. *Cognition, 20,* 191 - 208.

Baylis, G., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 451 - 470.

Behrmann, M., & Tipper, S. (1994). Object-based visual attention: Evidence from unilateral neglect. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance 15: Conscious and nonconscious processing and cognitive functioning* (pp. 351 - 375). Cambridge, MA: MIT Press.

Bertenthal, B. (1996). Origins and early development of perception, action, and representation. *Annual Review of Psychology, 47*, 431 - 459.
Bogartz, R. S., & Shinskey, J. L. (1998). On perception of a partially occluded object in 6-month-olds. *Cognitive Development, 13*, 141-163.
Bogartz, R. S., Shinskey, J. L., & Schilling, T. (1999). Object permanence in five-and-a-half month old infants? Manuscript in preparation.
Bogartz, R., Shinskey, J., & Speaker, C. (1997). Interpreting infant looking: The event set × event set design. *Developmental Psychology, 33*, 408 - 422.
Bower, T. (1974). *Development in infancy*. San Francisco: W. H. Freeman.
Burkell, J., & Pylyshyn, Z. W. (1997). Searching through selected subsets of visual displays: A test of the FINST indexing hypothesis. *Spatial Vision, 11*, 225 - 258.
Carey, S. (1995). Continuity and discontinuity in cognitive development. In E. Smith & D. Osherson (Eds.), *Thinking*, Vol. 3 of *An invitation to cognitive science*, 2nd Ed. (pp. 101 - 130). Cambridge, MA: MIT Press.
Chun, M., & Cavanagh, P. (1997). Seeing two as one: Linking apparent motion and repetition blindness. *Psychological Science, 8*, 74 - 79.
Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. *Nature, 371*, 791 - 793.
Downing, C., & Pinker, S. (1985). The spatial structure of visual attention. In M. Posner & O. S. M. Marin (Eds.), *Attention and Performance XI* (pp. 171 - 187). London: Erlbaum.
Driver, J., & Halligan, P. (1991). Can visual neglect operate in object-centered coordinates? An affirmative single case study. *Cognitive Neuropsychology, 8*, 475 - 494.
Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501 - 517.
Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48*, 269 - 297.
Egly, R., Driver, J., & Rafal, R. (1994). Shifting visual attention between objects and locations: Evidence for normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*, 161 - 177.
Enns, J., & Rensink, R. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review, 98*, 335 - 351.
Enns, J., & Rensink, R. (1998). Early completion of occluded objects. *Vision Research, 38*, 2489 - 2505.
Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics, 12*, 201 - 204.
Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40*, 225 - 240.
Goodale, M. (1995). The cortical organization of visual perception and visuomotor control. In S. Kosslyn & D. Osherson (Eds.), *Visual Cognition*, Vol. 2 of *An Invitation to Cognitive Science*, 2nd Ed (pp. 167 - 214). Cambridge, MA: MIT Press.
Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
He, Z., & Nakayama, K. (1994). Perceiving textures: Beyond filtering. *Vision Research, 34*, 151 - 162.
Heller, M. (1997). Gaps in perception. *Perception, 26*, 1481 - 1484.
Hirsch, E. (1982). *The concept of identity*. New York: Oxford University Press.
Hirsch, E. (1997). Basic objects: A reply to Xu. *Mind & Language, 12*, 406 - 412.
Intriligator, J. M. (1997). The spatial resolution of visual attention. Unpublished doctoral dissertation, Harvard University.
Johnson, A., & Aslin, R. (1995). Perception of object unity in 2-month-old infants. *Developmental Psychology, 31*, 739 - 745.
Johnson, A., & Aslin, R. (1996). Perception of object unity in young infants: The roles of motion, depth, and orientation. *Cognitive Development, 11*, 161 - 180.
Johnston, J., & Pashler, H. (1990). Close binding of identity and location in visual feature perception. *Journal of Experimental Psychology: Human Perception & Performance, 16*, 843 - 856.
Julesz, B. (1980). Spatial frequency channels in one-, two-, and three-dimensional vision: Variations on an auditory theme by Bekesy. In C. Harris (Ed.), *Visual coding and adaptability* (pp. 263 - 316). Hillsdale, NJ: Erlbaum.
Julesz, B. (1990). Early vision is bottom-up, except for focal attention. In *Cold Spring Harbor Symposia on Quantitative Biology - The Brain, 55*, 973 - 978. New York: cold Spring Harbor Laboratory Press.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29 - 61). New York: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 174 - 219.

Kanwisher, N. (1987). Repetition blindness: Type recognition without token individuation. *Cognition, 27*, 117 - 143.

Kanwisher, N., and Driver, J. (1992). Objects, attributes, and visual attention: Which, what, and where. *Current Directions in Psychological Science, 1*, 26 - 31.

Kellman, P. (1988). Theories of perception and research in perceptual development. In A. Yonas (Ed.), *Perceptual development in infancy* (pp. 267 - 281). Vol. 20 of the Minnesota Symposium on Child Psychology. Hillsdale, NJ: Erlbaum.

Kellman, P., & Spelke, E. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15*, 483 - 524.

Kellman, P., Spelke, E., & Short, K. (1986). Infant perception of object unity from translatory motion in depth and vertical translation. *Child Development, 57*, 72 - 86.

Kim, I., & Spelke, E. (1992). Infants' sensitivity to effects of gravity on object motion. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 385 - 393.

Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception, 13*, 287 - 305.

Leslie, A. M. (1986). Getting development off the ground: Modularity and the infant's perception of causality. In P. van Geert (Ed.), *Theory building in development* (pp. 405 - 437). North Holland: Amsterdam.

Leslie, A. M. (1988). The necessity of illusion: Perception and thought in infancy. In L. Weiskrantz (Ed.), *Thought without language* (pp. 185 - 210). Oxford: Oxford Science Publications.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. A. Hirschfield & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119 - 148). Cambridge, Cambridge University Press.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*, 265 - 288.

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

Maunsell, J. (1995). The brain's visual world: Representation of visual targets in cerebral cortex. *Science, 270*, 764 - 768.

Melkman, R., & Rabinovitch, L. (1998). Children's perception of continuous and discontinuous movement. *Developmental Psychology, 34*, 258 - 263.

Moore, M., Borton, R., & Darby, B. (1978). Visual tracking in young infants: Evidence for object identity or object permanence? *Journal of Experimental Child Psychology, 25*, 183 - 198.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104*, 686 - 713.

Nakayama, K., & Shimojo, S. (1990). Toward a neural understanding of visual surface representation. In *Cold Spring Harbor Symposia on Quantitative Biology - The Brain, 55*, 911 - 924. New York: Cold Spring Harbor Laboratory Press.

Nakayama, K., He, Z., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S. M. Kosslyn & D. Osherson (Eds.), *Visual Cognition*, Vol 2. of *An Invitation to Cognitive Science*, 2nd Ed (pp. 1 - 70). Cambridge, MA: MIT Press.

Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month old infants. *Cognition, 47*, 121 - 148.

Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology, 7*, 480 - 494.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Posner, M., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160 - 174.

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences, 3*, 111 - 169.

Pylyshyn, Z. W. (1984). *Computation and cognition: Towards a foundation for cognitive science.* Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial index model. *Cognition, 32*, 65 - 97.

Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50,* 363 - 384.
Pylyshyn, Z. W. (in press). Is vision continuous with cognition? The case for impenetrability of visual perception. *Behavioral and Brain Sciences*.
Pylyshyn, Z. W. (1999). *Seeing: An essay on vision and mind*. Manuscript in preparation.
Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179 - 197.
Pylyshyn, Z. W., Burkell, J., Fisher, B., Sears, C., Schmidt, W., & Trick, L. (1994). Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology, 48,* 260 - 283.
Quinlan, P. T. (1998). The recovery of identity and relative position from visual input: Further evidence for the independence of processing of what and where. *Perception & Psychophysics, 60*, 303 - 318.
Rafal, R. D. (1997). Balint syndrome. In T. Feinberg & M. Farah (Eds.), *Behavioral neurology and neuropsychology* (pp. 337 - 356). New York: McGraw-Hill.
Rock, I., & Gutman, D. (1981). The effect of inattention and form perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 275 - 285.
Sagi, D., & Julesz, B. (1985). Detection versus discrimination of visual orientation. *Perception, 14,* 619 - 628.
Schmidt, W. C., Fisher, B. D., & Pylyshyn, Z. W. (1998). Multiple-location access in vision: Evidence from illusory line motion. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 505 - 525.
Scholl, B. J. (1997). Cognitive development and cognitive architecture: Two senses of 'surprise'. Paper read at the 23rd annual meeting of the Society for Philosophy and Psychology, New York, 5/5/97.
Scholl, B. J. (1999). Varieties of 'top-down processing' in perception. Manuscript in preparation.
Scholl, B. J., & Pylyshyn, Z. W. (in press). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*
Sears, C. R., & Pylyshyn, Z. W. (in press). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology*.
Shimojo, S., & Nakayama, K. (1990). Amodal presence of partially occluded surfaces determines apparent motion. *Perception, 19*, 285 - 299.
Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A 'non-numerical' account. *Cognitive Development, 12*, 349 - 372.
Simon, T. J. (1998). Computational evidence for the foundations of numerical competence. *Developmental Science, 1*, 71 - 77.
Simon, T., Hespos, S., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development, 10,* 253 - 269.
Spelke, E. (1988a). The origins of physical knowledge. In L. Weiskrantz (Ed.), *Thought without language* (pp. 168 - 184). Oxford: Oxford Science Publications.
Spelke, E. (1988b). Where perceiving ends and thinking begins: The apprehension of objects in infancy. In A. Yonas (Ed.), *Perceptual development in infancy* (pp. 197 - 234). Hillsdale, NJ: Erlbaum.
Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition, 50*, 431 - 445.
Spelke, E. and Kestenbaum, R. (1986). Les origines du concept d'objet. *Psychologie Franciase, 31,* 67 - 72.
Spelke, E., Breinlinger, K., Macomber, J. and Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*, 605 - 632.
Spelke, E., Gutheil, G., & Van de Walle, G. (1995). The development of object perception. In S. Kosslyn & D. Osherson (Eds.), *Visual Cognition*, Vol. 2 of *An Invitation to Cognitive Science*, 2nd Ed (pp. 297 - 330). Cambridge, MA: MIT Press.
Spelke, E., Katz, G., Purcell, S., Erlich, S., & Breinlinger, K. (1994). Early knowledge of object motion: Continuity and inertia. *Cognition, 51*, 131 - 176.
Spelke, E., Kestenbaum, R., Simons, D. J. and Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology, 13,* 113 - 142.
Spelke, E., Vishton, P., & Von Hofsten, C. (1995). Object perception, object-directed action, and physical knowledge in infancy. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 165 - 179). Cambridge, MA: MIT Press.
Starkey, P., & Cooper, R. (1980). Perception of numbers by human infants. *Science, 210,* 1033 - 1035.
Starkey, P., Spelke, E., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36,* 97 - 128.

Strauss, M., & Curtis, R. (1981). Infant perception of numerosity. *Child Development, 52*, 1146 - 1153.
Streri, A., & Spelke, E. (1988). Haptic perception of objects in infancy. *Cognitive Psychology, 20*, 1 - 23.
Streri, A., Spelke, E., & Rameix, E. (1993). Modality-specific and amodal aspects of object perception in infancy: The case of active touch. *Cognition, 47*, 251 - 279.
Tanaka, Y., & Shimojo, S. (1996). Location vs. feature: Reaction time reveals dissociation between two visual functions. *Vision Research, 36*, 2125 - 2140.
Tipper, S., Brehaut, J., & Driver, J. (1990). Selection of moving and static object for the control of spatially directed action. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 492 - 504.
Tipper, S. P., Driver, J., & Weaver, B. (1991). Object-centered inhibition of return of visual attention. *The Quarterly Journal of Experimental Psychology, 43A*, 289 - 298.
Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97 - 136.
Tremoulet, P. (1998). Individuation and identification of physical objects: Evidence from human infants. Rutgers University Center for Cognitive Science Technical Report #41.
Trick, L., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 331 - 351.
Trick, L., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*, 80 - 102.
Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London, B203*, 405 -426.
Ungerleider, L, & Mishkin, M. (1984). Two cortical visual systems. In D. Ingle, M. Goodale, & R. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549 - 586). Cambridge, MA: MIT Press.
van de Walle, G., & Spelke, E. (1996). Spatiotemporal integration and object perception in infancy: Perceiving unity vs. form. *Child Development, 67*, 2621 - 2640.
van Loosbroek, E., & Smitsman, A. (1990). Visual perception of numerosity in infancy. *Developmental Psychology, 26*, 916 - 922.
Wellman, H., & Gelman, S. (in press). Knowledge acquisition in foundational domains. In D. Kuhn & R. Siegler (Eds.), *Cognition, perception, and language*. Vol. 2 of the *Handbook of child psychology* (5th Ed.). New York: Wiley.
Wiggins, D. (1980). *Sameness and substance*. Oxford, England: Basic Blackwell.
Wiggins, D. (1997). Sortal concepts: A reply to Xu. *Mind & Language, 12*, 413 - 421.
Wilson, J., & Robinson, J. (1986). The impossibly twisted Pulfrich pendulum. *Perception, 15*, 503 - 504.
Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research, 37*, 25 - 43.
Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749 - 750.
Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science, 7*, 164 - 169.
Xu, F. (1997). From Lot's wife to a pillar of salt: Evidence that physical object is a sortal concept. *Mind & Language, 12*, 365 - 392.
Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology, 30*, 111 - 153.
Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24*, 295 - 340.
Yantis, S. (1993). Stimulus-driven attentional capture. *Current Directions in Psychological Science, 2*, 156 - 161.
Yantis, S. (1995). Perceived continuity of occluded visual objects. *Psychological Science, 6*, 182 - 186.
Yantis, S., & Johnson, D. (1990). Mechanisms of attentional priority. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 812 - 825.
Yantis, S., & Jones, E. (1991). Mechanisms of attentional selection: Temporally-modulated priority tags. *Perception & Psychophysics, 50*, 166 - 178.