Mapping the mind

Domain specificity in cognition and culture

Edited by

LAWRENCE A. HIRSCHFELD

University of Michigan

SUSAN A. GELMAN

University of Michigan

CAMBRIDGE

UNIVERSITY PRESS 1994

5 ToMM, ToBy, and Agency: Core architecture and domain specificity

Alan M. Leslie

Our understanding of Agency is, in part, the result of domain-specific learning. The nature of this domain-specific learning needs to be understood in relation to the organization of information processing in the infant. As a result of adaptive evolution, the infant is a specialized processor of information with an architecture that (in part) reflects properties of the world. On this assumption, it should be possible to establish links between properties of the world, processing subsystems specialized for tracking those properties, and domains of knowledge. It is argued in the case of Agency that three main classes of world properties are reflected in three corresponding processing subsystems producing three distinct levels of knowledge. These three related triples are, respectively, *mechanical* Agency, *actional* Agency, and *attitudinal* Agency. Each of these three linked property classes, processing subsystems, and knowledge levels are discussed in turn but the focus will be mainly on mechanical Agency. In developing these ideas this discussion deals more generally with the nature of early mechanical understanding and its relation to conceptual development. A number of the ideas put forward in Leslie (1988) are revised and extended.

One lesson of cognitive science is that different types of knowledge often have different locations within the global organization of human information processing. In development, different types of commonsense knowledge may originate from different locations in core cognitive architecture. Early mechanical understanding and the notion of Agency can be studied within such a framework.

Parts of this chapter were written while I was visiting the Psychology Department at UCLA. I greatly benefited from discussions and conversations with a number of people there including Randy Gallistel, Nancy Kanwisher, and Liz Spelke who passed through, but most particularly with Rochel Gelman who gets my special thanks. I am also grateful to Geoff Hall, John Morton, and Jean Mandler for helpful comments on an earlier draft.

Domain specificity

The trouble with the notion of domain specificity is that there could turn out to be too many domains. For example, we may be tempted to describe knowledge of chess as a domain distinct from car driving, or we may wonder whether chess and checkers form one domain or two. In the study of development, it is customary to stick quite closely to the description of the child's conscious knowledge or lack thereof, often to the exclusion of all else. This one-dimensional approach to cognition encourages the proliferation of "domains." However, to study the cognitive mechanisms that produce development it is necessary to do more than enter the remote lands of childhood and carry back reports of quaint beliefs and astounding ignorance. To the extent that there are mechanisms of domain-specific development, then a deeper notion of domain is possible − one that is less software dependent, less profligate, and more revealing of the design of human cognition. This kind of domain specificity reflects the specialization of mechanisms in core cognitive architecture.

By core cognitive architecture I mean those human information processing systems that form the basis for cognitive development rather than its outcome (Leslie, 1988). Understanding this core is the primary aim of all theories of cognitive development. One view of the core is that it is essentially homogeneous and that any differentiation of its architecture is the product of development. The general all-purpose learning device of classical associationism is an elegant and influential example of this view. An alternative view of the core is that it contains heterogeneous, task-specialized subsystems. Vision is an obvious example of a specialized subsystem with a specialized internal structure. The language faculty is another. I propose and discuss a third specialized system that interfaces input and central processes and that structures the development of conceptual knowledge.

Types of domain-specific mechanisms

Different kinds of specialized core devices can underlie domain specificity. Some mechanisms may perform specialized tasks, not because they are particularly special on the inside − they have no special processes nor a prestructured representational system − but because they occupy a special position within the overall processing organization. The positioning of the device guarantees it will receive input from a particular class of object in the world and that it will end up representing a certain kind of domain-specific information. Mechanisms for face recognition may be an example of this. Faces seem to be processed by a device that employs general, non-specialized pattern processing (Diamond & Carey, 1986; Ellis & Young, 1989; Tanaka & Farah, 1991) but that receives only faces as input. Johnson and Morton (1991) propose that in the first few weeks of life the face recognition device, which they call "CONLERN," receives restricted input because of another device they call "CONSPEC." Unlike CONLERN, CONSPEC *is* internally specialized − it contains a rudimentary template of a face that serves to direct attention to faces. On this story, CONLERN becomes a face recognition device because it receives only face inputs.

A different kind of core domain-specific device is exemplified by a mechanism that acquires the syntactic structure of the natural language to which it is exposed. The study of language acquisition suggests that such a device not only occupies a special position in overall organization and is fed a special class of input, but also processes that input in a relatively specialized way, and, in so doing, employs a specialized representational system (e.g., Chomsky, 1975, 1986; Pinker, 1989).

The language faculty is probably not the only member of a class of core domains concerned with knowledge of formal systems. Formal core domains plausibly include *number* and perhaps *music,* as well as grammar. In formal domains, structural relations are key and there is no organizing role for the notion of cause and effect. By contrast, the notion of cause and effect is the central organizing principle in the core domains of object mechanics and "theory of mind." Like Carey (1985), I believe that these two core domains comprise the major part of our initial capacity for causal conceptual knowledge. These causal domains are my focus here and I shall argue that specialized core devices drive their development.

An overview of the core architecture of Agency

I want to examine a relatively neglected topic in studies of development, namely, the notion of Agency. I capitalize the first letter because, in some uses, an agent is simply a cause. The notion of an Agent, however, is more restricted and, in the first instance, applies to a certain class of object. Before discussing the properties that distinguish this class of object and define the notion of Agent, I want to draw attention to a distinction that is probably important but that I shall not pursue here. This is the distinction between Agent and animate object. Most objects that are Agents are animate and certainly all the objects that ever, in the course of evolution, contributed to the adaptation of our cognitive systems for dealing with Agency were animate. Nevertheless, I assume that the notion of animateness is external to Agency and proprietary to the biological domain. I can leave open the question whether or not biological knowledge constitutes a domain in core architecture (but see Keil, this volume), and if so what type.

I propose that the notion of Agency emerges from domain-specific learning and reflects properties of core architecture. In exploring the relationship between core architecture and our ability to understand the behavior of Agents, I will postulate two processing devices: First I shall discuss **ToBy** (Theory of Body mechanism), the seat of the infant's theory of physical

Table 5.1. *Core architecture for the cognition of Agency*

A Tripartite Theory of Agency		
Real World Properties of Agents	Processing Device	Levels of Understanding or "theories"
mechanical	**ToBy**	"Agents and Objects"
actional	**ToMM**(system$_1$)	"Agents and Action"
cognitive	**ToMM** (system$_2$)	"Agents and Attitudes"

bodies; later I shall discuss **ToMM** (the Theory of Mind Mechanism), the seat of the child's "theory of mind."[1]

I propose that understanding Agency is not achieved by a single conceptual system nor by a single processing system. Instead, it involves three distinct, hierarchically arranged processing components or modules. The three components correspond to or achieve three distinct levels of understanding or "theories" of Agency.

The first component, which I call **ToBy,** embodies the infant's theory of physical objects. **ToBy** is concerned with Agents in a mechanical sense - that is, with the mechanical properties of Agents. Distinguishing Agents from other physical bodies that are not Agents and describing their mechanical interactions are important functions of **ToBy.**

The next two components and their corresponding levels of understanding are concerned with the "intentional" properties of Agents. Although the movements and states of mere objects are simply features *of* the world, the movements and states of Agents, as well as being *of* the world, are also *about* the world (see, e.g., Dennett & Haugeland, 1987, for this notion of "aboutness"). The aboutness or intentionality of action and cognition are dealt with at the next two levels respectively. Together these two components, which deal with the intentionality of Agents, make up the device I call **ToMM.** Unlike **ToBy, ToMM** is exclusively Agent-centered. The first subcomponent of **ToMM** *("system$_1$")* is concerned with Agents and the goal-directed actions they produce. This second level theory of Agency can be called "Agents and Action."

The third and final level of the hierarchy is concerned with the mental states of Agents and their role in producing behavior. At this level, Agents are represented as holding attitudes to the truth of propositions − attitudes such as *wanting, believing,* and *pretending* that *p,* where *p* is a proposition of some kind. This third level theory of Agency can be called "Agents and Attitudes." Although the first level of Agency is part of the infant's "theory of physical bodies," the latter two levels are part of the infant's theory of mind. This tripartite theory of Agency is summarized in Table 5.1.

This hierarchy is not a series of stages of development in the traditional sense. Instead, each level corresponds to a separate subsystem. Each component or subsystem constitutes a learning device with a specialized agenda for information acquisition, a specific way of organizing or describing the inputs it receives, a specific location within the larger architecture, and therefore with particular relationships to other components. The development of each subsystem can proceed in parallel and unfold according to its own distinct character and inputs. Naturally, each subsystem could *begin* its development in sequence, determined in part by the maturational status of the appropriate neural circuits and in part by the availability and quality of inputs. Thereafter, each subsystem can develop in parallel.

My discussion concentrates upon the first of these levels, mechanical Agency, and deals with the other two levels only to mark off what else I think comprises the core notion of Agency that is not dealt with at the first level. I begin then with the emergence of a processing mechanism (probably somewhere around 3 or 4 months of age) that equips the infant to attend to the mechanical properties of objects and events.

ToBy: A mechanics module

Piaget's (1953, 1955) view of the infant's developing knowledge of the physical world was tied to his view of core architecture and thus to his view of infant learning mechanisms. As in classical associationism, core architecture is assumed to be homogeneous and unstructured. The core, according to Piaget and to classical associationism, consists of two things: an ability to represent microfeatural sensations, and a set of completely general learning procedures. For Piaget, the learning procedures operate iteratively over the microfeatures (and neonatal reflexes) to build schemas. Differentiation or specialization of architecture is purely the result of psychological development and never initially its cause. The homogeneity assumption dictated the gradual and uniform cognitive development that Piaget thought he observed in examining structured action in infancy. Thus, Piaget believed that it was not until the end of infancy, toward the close of the second year, that infants construe the physical world as a rigid three-dimensional space containing stable, enduring objects whose behavior is regulated by causality. Before then, the infant processes the world as a disorganized display, whose chaos gradually gives way to familiarity, but whose meaning depends entirely on the infant's present activity. The world beyond subjective activity is, for the infant, simply a void.

The picture, which we read in Piaget, of the infant's painfully slow construction of an objective world seems now to reflect more on the limitations of the infant's capacity for planned, structured activity, for example, in manual search tasks (Diamond, 1988), and less on the infant's capacity to represent stable bodies with mechanical properties. A different picture of infantile representation emerges when visual attention measures are used to probe infant cognition. Measuring the infant's attention to events apparently taps a different psychological system than that which governs manual search behavior. Piaget's view overlooks even this gross degree of modularity. According to the new measures, infants, a few months after birth, apprehend cohesive, bounded, spatiotemporally continuous objects (see Spelke, 1990, for review), attribute specifically causal properties to collisions between objects (see Leslie, 1988, for review), and model some of the properties of hidden objects (see Baillargeon, 1991b, for review). Apparently infant competence even extends to construing the hidden mechanism in Piaget's classic invisible displacement event (Leslie & DasGupta, 1991). The infant's processing of the physical world appears to organize rapidly around a core structure representing the arrangement of cohesive, solid, three-dimensional objects embedded in a system of mechanical relations, such as pushing, blocking, and support.

These findings, in my view, inform us about a specialized learning mechanism adapted to create conceptual knowledge of the physical world, and to do so at an early period in development when general knowledge and general problem-solving abilities are quite minimal. Leslie and Keeble (1987) suggested that modular organization provides a way to work around the inevitably limited general capacities and knowledge characteristic of the preschool period. Core modularity provides a way to ensure rapid and uniform knowledge acquisition in domains that have adaptive significance to our species. Such acquisition secures, in turn, the early success of informative communication and thus the ability to take part in and benefit from the cultural transmission of knowledge. Although the best evidence for modular or compartmentalized processing comes from the study of perceptual input systems (Fodor, 1983; Marr, 1982), developmental benefits would also accrue from componential organization at more central levels of processing.

Alternatively, one could adopt a view that stresses the similarities and continuities between commonsense theories and theories that are the products of science or other forms of deliberate, reflective thinking. One version of this view places the conscious manipulation of concepts at center stage. Although imagining the child consciously solving a problem gives one a reassuring sense of having understood a piece of development, in fact this assurance is entirely illusory. Conscious thought is no better understood than the mass of unconscious processes upon which it depends.

If one takes a "child-as-scientist" view and pictures the child as an ordinary everyday scientist, working hard to contribute additional phenomena to an existing theoretical framework, then of course one addresses the nature of that original framework. The ideas set forth here can be aligned with such a picture. If instead one recoils from initial structure and starts from an unconstrained, general core architecture, the child-as-scientist metaphor changes in a critical way. Now the metaphor requires one to picture the child as a *great* scientist, begetter of conceptual revolution and radical theory shift.

This child-scientist produces her conceptual revolutions without the benefit of formal instruction, does so regularly, and in several different domains simultaneously. Her astonishingly successful and prolific early career is diminished only by the fact that all other children make essentially this same progress too, in essentially the same way, without effort, and, by and large, independently of IQ.

In view of these facts, it seems likely that some theories bootstrap thanks to specialized devices in core architecture. These processes can establish the core knowledge and ability on which pedagogy itself depends. **ToBy,** I will argue, is responsible for the early and rapid emergence of knowledge of physical bodies. Following Marr (1982), we may ask what is the goal of the information processing task this device carries out. **ToBy's** goal, in general, is to arrive at a description of the world in terms of the mechanical constitution of physical bodies and the events they enter into. There are two main parts to achieving this goal. The first part, following Spelke, is to find the stable three-dimensional objects in the world. The second part, following Gelman and Talmy, is to find the "sources of energy" that produce their motions (Gelman, 1990) or, more generally, the distribution or dynamics of forces in a scenario (Talmy, 1988).

ToBy and FORCE

ToBy is concerned with the mechanical properties of objects and events. My first assumption is that **ToBy** employs a primitive notion that I shall call FORCE. I reserve the word *energy* for talking about mechanical forces *in the world.* For the cognitive correlate of energy, I use the term FORCE. FORCE is meant to be a "primitive" − a commonsense notion introduced by a modular system that has resulted from evolutionary adaptation. FORCE is not the scientific term developed culturally. I want to mark this distinction in my terminology. The fact that in the world object motions are invariantly the result of energy distribution is what makes it advantageous that the psychological system has a way of attending to and representing sources of energy. This does not mean that **ToBy** possesses a complete or even very deep system for thinking about the physical world, when viewed from the point of view of modern or even of, say, Aristotelian scientific theory. Others have discussed psychological notions similar to FORCE (for example, some aspects of Shultz's [1982] "generative transmission," and Anderson's [1990] force models), the closest being Talmy (1988).

The employment of the notion of FORCE is principally what makes it the case that **ToBy** is concerned with *mechanics.* It also dictates that **ToBy** be interested in three-dimensional bodies. This follows from the fact that, in the world, only three-dimensional bodies have mechanical properties. More terminology: It seems likely that some version of the distinction reflected in language by the distinction between "mass" noun (e.g., *butter, water)* and "count" noun (e.g., *dog, table)* is available to **ToBy**. For count objects, particular individuals can be identified and counted (three tables, many dogs), whereas masses cannot ("much water" but not "three waters") unless a count object organizes the mass (e.g., "three glasses of water"). I shall try to use the term *object* where a specifically count object is intended and "body" when surfaces, masses, *and* objects are intended.

ToBy is concerned with three-dimensional objects as the principal bearers, transmitters, and recipients of FORCE. Although entities such as letters or moving patches of light on a surface may be treated as objects by perceptual processes that trace their identity (Treisman, 1988), they fail to be bodies with mechanical properties.[2]

A dual route to mechanical analysis

ToBy is not concerned with entities that, like patches of light or letters on a page, lack mechanical properties. In contrast, three-dimensional objects can hardly fail to have mechanical properties. Spelke (1988, 1990, this volume) has argued persuasively that infants share with adults certain core aspects of the notion of physical object as bounded, cohesive, solid, three-dimensional bodies whose existence is spatiotemporally continuous. Infants also apparently regard objects as opposed to masses as countable (S. Carey, personal communication). Such a notion of physical object, which I attribute to **ToBy** and its specialized representational system, requires the availability of mechanical description. Therefore, I argue for the following proposition. The concern with picking out and tracking physical objects − a concern that unites adult and infant − is in one guise a concern with the mechanical structure of the world. The key player in this mechanical world is the three-dimensional object.

I wish that my account could have been as simple as the previous paragraph promises. However, mechanical events typically involve an element of motion. It is a fact about the world that the motion of three-dimensional objects is the result of the distribution of energy. Motion is therefore a source of information about FORCE. It is a fact about the visual system that it processes motion without reference to whether or not the bearer of motion is a three-dimensional object. Vision deals with the complexities of recognizing the form of three-dimensional objects too late in the stream of processing to be of use in the analysis of motion. Motion analysis takes place independently of the processing of luminosity, color, and texture, and earlier than the analysis of shading and the occlusion of surfaces (VanEssen, 1985; Livingstone & Hubel, 1988). These two facts, one about the world, the other about visual processing, suggest there will be a second route to mechanical analysis, a route that is independent of three-dimensional objects. Supporting evidence for this would be the existence of special visual sensitivities to patterns of motion that are highly informative with respect to FORCE dynamics. Some

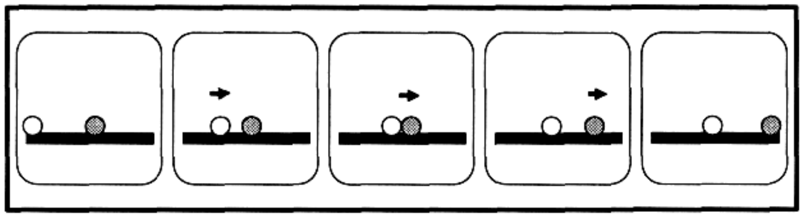

Figure 5.1. The "billiard ball" launching event.

of these patterns were discovered by Michotte (1963), chief among them being the "billiard ball launching" event.

In a launching event, one object moves toward another stationary object, strikes it sending that object off, while the first object now becomes stationary at the point of impact (see Figure 5.1). Michotte discovered that adults had an immediate impression of cause and effect when viewing this configuration. Two things are striking about his discovery. First, events exactly like this almost never occur in the real world. (For example, such a pattern with objects of equal mass occurs only in 180-degree collisions if the objects have perfect elasticity.) Second, the impression of causality occurs despite the fact that the objects involved are only pencil marks on paper or patches of light on a wall. The observers know perfectly well that there is no real mechanical connection between the motions of the insubstantial, two-dimensional entities. Nevertheless, the causal impression is quite incorrigible. This led Michotte to argue, correctly in my opinion, for the radical notion that adults are subject to a perceptual *illusion* of causality.

In a series of studies, I showed that, by 6½ months and possibly earlier, infants too are subject to the launching illusion (Leslie, 1982, 1984b; Leslie & Keeble, 1987; reviewed in Leslie, 1988). I proposed that the perception of cause and effect in infant and in adult were linked through a modular component of motion analysis that serves to kick start development in infants and has the side-effect of creating a causal illusion in adults. I shall call this component the *Michotte module.*[3]

The *dual route* to mechanical analysis suggests that **ToBy** has two principal inputs from vision: one from a three-dimensional object recognition device, and one from motion analysis systems, including the Michotte module. Following Warrington (Warrington & James, 1986; Warrington & Taylor, 1978) and Marr (Marr, 1982; Marr & Nishihara, 1978) and others, I make the following assumption. The representation of the shape of an object involves a process distinct from the representation of other kinds of information about objects such as their use or function. In brain-damaged patients either kind of information may be impaired independently of the other. The visual three-dimensional object recognition device is concerned purely with the "geometry" of objects, recognizing them by matching to a three-dimensional shape model stored in a catalogue.[4] Specifically, visual object recognition is not concerned with the mechanical properties of the object, and therefore is not concerned with whether the object is cohesive, substantial, mechanically bounded, or numerically identical over time. These aspects of the "object concept" are the concern of **ToBy.** I shall refer to the output of the visual object recognition module as the *purely visual object* to underline the distinction between an object-recognized-by-shape and an object-constituted-mechanically.

In summary: Together with information on surface layout, **ToBy** takes, as input, descriptions that make explicit the geometry of the objects contained in a scene, their arrangement and their motions, and onto such descriptions paints the mechanical properties of the scenario. In doing this, **ToBy** interprets the motions, arrangements, and geometry of the objects in terms of the sources and fates − the dynamics − of FORCE.

Purely visual causality versus mechanical causality

The two principal visual inputs to **ToBy** determine a dual route to generating mechanical descriptions. **ToBy** paints mechanical properties onto purely visual objects (and more generally, onto bodies and surfaces). However, as we have seen, certain patterns of motion also attract FORCE descriptions even though no three-dimensional objects are involved.

The Michotte module, although it renders a perception of cause and effect, does not produce a FORCE description. Instead, the cause and effect of the Michotte module is a disembodied or purely visual cause and effect. This assumption is close to what Michotte himself believed. Michotte rejected the notion that his launching effect depended upon the perception of force (Michotte & Thines, 1963) and instead, related it to what he called "ampliation of the movement." This was a phenomenological notion that seems to be best interpreted as a purely visual − that is, spatiotemporal − extension of the movement of the first object in the second.

The distinction between Marr's purely visual three-dimensional object, on the one hand, and the mechanical object (the cohesive, solid, bearer of FORCE), on the other hand, is echoed in the distinction between Michotte's "purely visual" causality and mechanical causality based on the dynamics of FORCE. These distinctions probably reflect a more general architectural distinction. It seems characteristic of vision that what it makes explicit are the *spatial* properties of surfaces, objects, and motions. Marr called this "the quintessential fact of human vision - that it tells us about shape and space and spatial arrangement" (Marr, 1982: 36). Any information processing system has to represent information. Any system of representation brings particular kinds of entity and particular kinds of information to the fore − makes them "explicit" to use Marr's term − whereas other kinds of information

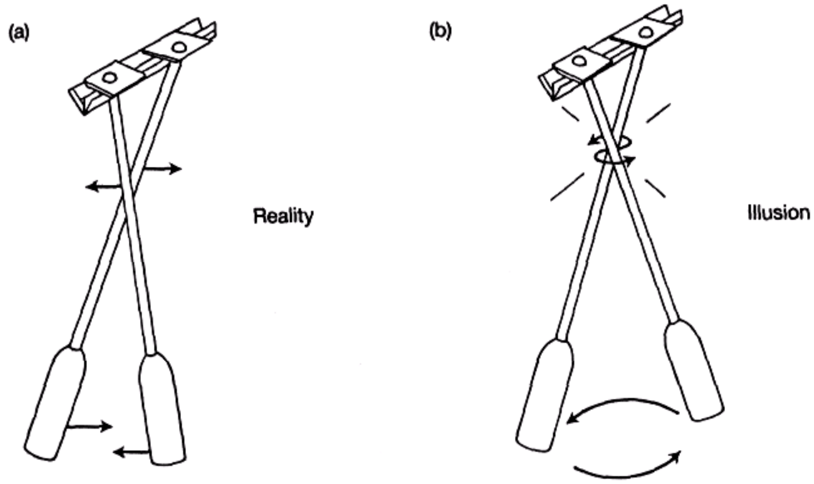

Figure 5.2. Pulfrich double pendulum illusion. Wearing a neutral density filter over one eye but viewing the swinging rods with both eyes creates a stereoscopic depth illusion. The result is that subjects see the solid rods passing through one another (after Leslie, 1988).

are left implicit or are pushed into the background. Approaching vision as an information processing system, it is apparent that what is made explicit by visual processing is spatial information, including, of course, spatial arrangement over time (i.e., motion).

I want to claim that the concern of vision for spatial description excludes the description of mechanical properties. Mechanical properties are left implicit "in the background." For example, in the stereoscopic illusion produced by a Pulfrich double pendulum (see Figure 5.2), the solid rods of the pendulum appear to pass through one another (Leslie, 1988). Vision is not constrained to suppress the illusory trajectories of the rods simply to prevent the mechanical anomaly of the appearance of passing through, though other *spatial* anomalies can have this suppressing effect (Leslie, unpublished). It appears that a mechanical solidity constraint is not employed in visual processing. Furthermore, the ease of forming a visual image of two solid objects passing through one another suggests that neither vision nor visual imagination employs a solidity constraint. Despite this, infants only a few months old are surprised to view a scenario in which a hidden object appears to violate the solidity constraint (Baillargeon, 1986; Baillargeon, Spelke, & Wasserman, 1985). Despite the indifference of vision and visual imagination, the solidity constraint shows up in naive mechanical reasoning. The reason for this intriguing pattern, I suggest, is that mechanical constraints are not the province of vision − neither of visual imagination nor of visual experience − but of **ToBy.**

Space and mechanics are not the same (or may the FORCE be with you)

Although spatiotemporal patterns are highly confounded with *contact* mechanics, the two are not the same. Unfortunately, the correlation can encourage the idea that mechanical notions reduce to mere spatiotemporal patterns and that mechanics is not fundamental to our understanding of the world. This is, of course, the starting point of classical empiricism (e.g., Hume, 1740) and the idea finds many echoes elsewhere. Michotte provides examples: His distinction between "mere displacement" and "movement" (Michotte & Thines, 1963) cries out for interpretation in terms of whether the moving entity is seen as the dominant bearer of FORCE (= "movement") or not (= "displacement"), as in the case of one object transporting another. Likewise, as I will argue below, "ampliation" of the movement in a launching event is most naturally interpreted as the transmission of FORCE from one object to another through contact.

More recently, Mandler (1992) has put forward an account of infant competence that assumes that mechanical notions can be reduced to spatiotemporal patterns.[5] Mandler argues that conceptual development proceeds out of a perceptual analysis of the spatiotemporal properties of objects and events. This perceptual analysis yields a kind of representation that she calls "image schemas." Image schemas are analogue spatial representations and provide the earliest "concepts." The core architecture for conceptual development, in Mandler's view, employs a purely analogue, "non-propositional" format of representation. Specifically, image schemas are said to underlie the infant's notions of causation (launching), containment, and agency. The image schemas for these events are illustrated in Figure 5.3.

The reader should guard against the temptation to read mechanical meaning into the spatiotemporal patterns depicted by the image schemas in Figure 5.3. It is more natural for us to think about launching, containment, and agency events as mechanical events rather than as purely spatiotemporal patterns. Because of this, we may inadvertently think about the representation (the image schema) in terms of what it *refers to* rather than in terms of how it represents what it refers to. Of course, what the schema for, say, launching refers to is, as a matter of fact about the world, a mechanical event. But that is not how Mandler wants us to interpret her notion. Image schemas are defined by Mandler as representations that make explicit *spatial* (that is, spatiotemporal) information, not mechanical information. Indeed, this is crucial to her claim that analogue, spatial image schemas *alone* provide sufficient grounding for conceptual development.

However, the evidence for infants' grasp of launching (and as we shall see later, also for their grasp of containment and of Agency) shows that they understand the mechanical, and not just the spatiotemporal, properties of such events. Leslie and Keeble (1987) were able to demonstrate a causal

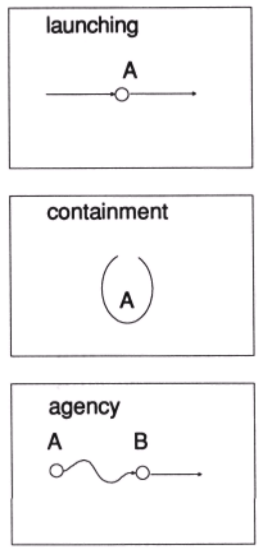

Figure 5.3. Mandler's "image schemas" for launching, containment, and agency (after Mandler, 1992).

illusion of launching in 6-month-olds precisely because they found a way to vary the mechanical properties of the test stimuli *while equating the changes in spatiotemporal properties.* This allowed them the conclusion that infants had reacted to something above and beyond spatiotemporal properties, namely, the causal (mechanical) properties of launching. More generally, we have to ask ourselves why infants analyze out the particular spatiotemporal patterns Mandler identifies. There are myriad other spatiotemporal patterns that could have been latched onto instead. The answer is, I believe, because these are the spatiotemporal patterns upon which the infant's theory of mechanics bestows immediate significance.

Image schemas, then, make explicit the wrong kind of information, if they are to play the role in conceptual development that Mandler assigns to them. We might enlarge on Mandler's theory and propose that **ToBy** operates over image schema-like representations, enriching them by painting on mechanical information. But this will create a different class of internal representation: The schemas will no longer be simple analogue images, but will become instead *mechanical diagrams.* Diagrams are partly spatial analogue representations, to be sure, but they are also partly symbolic. And mechanical diagrams, whether appearing statically upon the printed page and being interpreted by a reader or unfolding over time in a baby's head and being processed by the baby's cognitive system, depend for their efficacy upon the availability of mechanical notions. Such notions must rest upon at least a rudimentary theory of mechanics.

Mandler interprets work by Choi and Bowerman (1991) on "spatial" verb systems in support of her thesis about the primacy of spatial representation in development. But verb systems too encode mechanical notions; they pick out the spatial relations that have a significant mechanical interpretation. To take but one example from Choi and Bowerman, cited by Mandler, Korean (unlike English) uses a systematic marking for whether two objects or parts of an object fit together tightly or fit together loosely. There are, of course, geometric correlates of fitting tightly and fitting loosely − roughly, the closeness or remoteness of contact. But just as in the hand-as-agent experiments I will discuss below, the spatial relationship is relevant precisely because of the mechanical properties it signals. The significance of this spatial relationship − why it is attended to, why it is informative and worth encoding in a verb system − is clearly mechanical: How much force is required to make or break the contact? Will one object provide support for the other? Will the fitting just drop out if I don't hold it in there? And so on. The significant generalization to make from these cases is not primarily in terms of spatial representation but in terms of *the interpretation the spatial relations receive in a FORCE theory of contact mechanics.* This is the critical information grasped by early learners of verb systems. I believe much the same goes for most of the examples from perception in prelinguistic infants, which Mandler discusses, including launching, Agency, and containment.[6]

Part of the motivation Mandler has for proposing the notion of image schema as the basis for conceptual development and language acquisition is to avoid attributing "propositional" (i.e., predicate-argument) representation to the infant. However, the plausibility of this solution is diminished if mechanically interpreted FORCE diagrams are a minimum requirement for internally representing the notions Mandler targets. For example, this would require enriching analogue spatiotemporal representations by adding on symbol structures. Actually, predicate-argument structures are eminently suited to representing FORCE dynamical notions. Representing mechanical *roles* and mechanical *relations* raises many of the "binding" problems that predicate-argument structures are so good at solving. Indeed, mechanical roles and relations seem to be reflected directly in much of the verb-argument structure of natural language (Talmy, 1988) that appears largely to be specialized for this task, perhaps as a result of the adaptive coevolution of cognitive and linguistic abilities (Pinker, 1989; Pinker & Bloom, 1990).

Motion and motive FORCE

I now need to say something about how **ToBy** equips the infant to understand and learn about the moving world of physical objects. I shall consider this question in connection with three classes of events. I start with one of the simplest of all events.

Nothing terrestrial moves without having a source of energy. As Gelman

(1990) points out, this underwrites a basic or "first principle" of attending to motions, namely: Attend to the source of energy. Following this principle, there are only two possibilities when observing an object that begins to move. The first is that it was made to move by something else (which you may or may not be able to see) in which case its energy came from some other object. Or the object has an internal source of energy, in which case it is an Agent. So, in painting on a FORCE description, ToBy attends to sources of energy. The more an object changes motion state by itself and not as a result of external impact, the more evidence it provides, the more likely it is, that it is an Agent.[7] Notice that none of this follows simply from seeing the spatiotemporal characteristics of the motions per se, but from *interpreting* those characteristics in terms of FORCE source. As always, spatiotemporal properties are confounded with mechanical ones. Without access to the mechanical interpretation, however, none of the inferences concerning cause or Agency can be drawn and we are stuck, like David Hume, with the "impressions of our senses," namely, meaningless spatiotemporal patterns. As we shall see in the following third example, having access to this very simple FORCE dynamical interpretation (internal/external source) allows the infant to recognize a more complex class of events, namely, interactions.

My second example is the launching through collision event that has already been mentioned. Hume's celebrated analysis of causation began with this event. He considered one billiard ball colliding with and launching another to be "as perfect an instance of the relation of a cause and effect as any which we know..." (Hume, 1740). But he was quite unable to say *why* it should seem so perfect. Lacking any place for core mechanical understanding in his framework, Hume had to rely on the statistics of association. But launching has special properties for infants as well as for aficionados of the billiard table. From the point of view of ToBy's contact mechanical theory, launching is the simplest and most complete instance of the transmission of FORCE.

Under a FORCE interpretation, the two objects in launching are assigned different and imbalanced mechanical roles, one as pusher (transmitter of FORCE), the other as pushed (recipient of FORCE). As I noted earlier, when Leslie and Keeble (1987) found a way of unconfounding spatiotemporal properties, 6-month-olds showed they were sensitive to these mechanical roles. One group of infants watched a film of a causal looking launching event, while another group was shown a variation on launching. In this variation, the causal impression is destroyed by introducing a short delay (half second) between the impact of the first object and the reaction of the second. Adults do not think this delayed event looks like pushing and being pushed. Leslie and Keeble habituated one group of infants to a causal sequence and another group to a noncausal sequence. They then tested the two groups by showing them exactly the same film they had respectively seen before, except that the film projector now ran in reverse. The spatiotemporal properties of the sequences thus change: The spatial direction of motion changes (e.g., from left to right) and the temporal order of motion changes (e.g., red object moves first, green object second changes to green first, red second). But these spatiotemporal changes occur equally for *both* groups. Moreover, whatever degree of spatiotemporal continuity there was in the pattern of motion the infant was habituated to is still there unchanged in the test presentation. It must be: It is the same piece of film they saw before.

However, specifically in the case of launching, reversal produces a change in *mechanical roles* and this change introduces a difference between the two groups. In the causal event, reversal swaps the roles of pusher and pushed between the objects. The pusher becomes the pushed. In the noncausal event, these roles are absent and so they cannot be reversed. Thus, if infants appreciate the mechanical structure of launching, they will recover attention more to its reversal than to the reversal of a variant that lacks that same mechanical structure. That is what Leslie and Keeble found. Six-month-olds interpret spatiotemporal patterns in terms of mechanical structure.

Recently, Baillargeon (1991a) has found that young infants have expectations regarding the amount of FORCE a moving object imparts to a stationary one on impact, with greater FORCE being delivered by larger objects and resulting in greater distances traveled.

The third and final class of event takes us back to Agency and the interaction between an Agent object and a non-Agent object, as when a hand picks up a doll. Leslie (1982) habituated 5- and 7-month-old infants to such an event. Half the infants were then tested on the same event but with a number of visible changes including a different hand, different direction of movement, different pace, and so on. These infants did not recover interest to these changes. The other group of infants saw an event in which the only change was a small gap introduced between hand and doll so that when the hand again picked up the doll it looked "as if by magic." These infants did recover interest. This was followed up by Leslie (1984a) who showed that 6-month-olds were surprised at the lack of contact only during the pick-up and not when the hand and doll were stationary. Most important, they were surprised only when a hand was involved. If an event with the same spatiotemporal pattern of motion was shown that did not include a hand (the doll was "picked up" by a Styrofoam block), the infants were unconcerned about the spatial contact (see Figure 5.4). These results show that infants were concerned about spatial contact only when they thought a mechanical relationship was involved. And they thought a mechanical relation was involved only when a hand was at work.

Consider: An infant can often observe hands moving "on their own." This gives **ToBy** strong evidence that hands have an internal source of power. Hands are Agents. When the infant then sees a hand moving simultaneously and in contact with another object, *in spite of the precise spatiotemporal reciprocity of the two motions,* **ToBy** interprets the event as having a particular

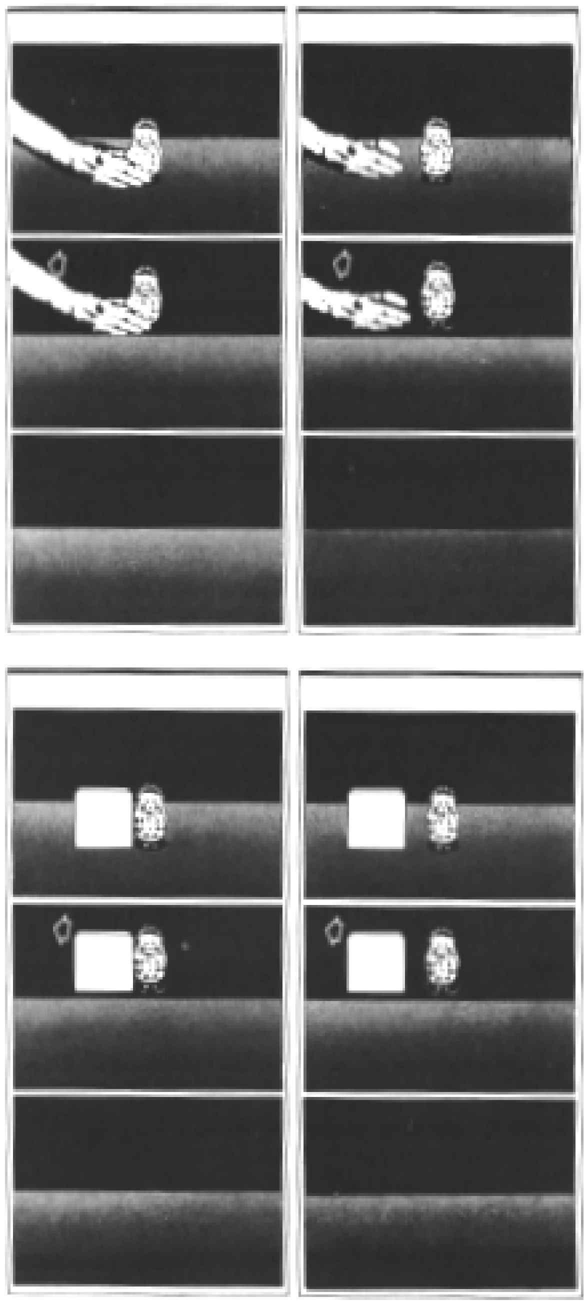

Figure 5.4. Infants perceive hands as Agents. This differentially influences their interpretation of the spatial relation in the top and bottom pairs of events (from Leslie, 1984a).

mechanical direction and involving unequal mechanical roles. The hand, as the source of FORCE in the event, must pick up and pull the doll rather than the other way around (where the doll pushes the hand backward). This is (to us) such an obvious way to interpret this event that it is worth laboring the point. In the past, the infant has observed hands and dolls individually exhibiting different kinds of spatiotemporal motion. But it does not follow merely from this fact that, in the future, when the hand and doll move together, the infant must see the hand as the causally active object, that is, as doing the picking up. But this *does* follow if the infant interprets the different prior spatiotemporal patterns as revealing different mechanical properties. Mandler and I agree that these studies show the infant's grasp of Agency. However, I sharply disagree with her proposal that this is achieved by a purely spatiotemporal analysis. The theory of ToBy provides, in my view, a fuller account of the mechanism whereby infants learn to identify and comprehend the behavior of Agents.[8]

Two questions for ToBy

In Piaget's (1955) theory of the origins of causal notions there are said to be two initial aspects to causality that gradually diverge in development. One is simply the tendency to associate together regular sequences of events. The other, which Piaget calls "efficacy," is the infant's supposed awareness of sensations of effort and desire accompanying action. I draw attention to efficacy because it might at first sight be confused with the following, in my opinion, sounder idea. Because of its position in global architecture, **ToBy** can use evidence provided by senses that are less wholly spatial than vision. In particular, ToBy can take advantage of information from kinesthetic, haptic, and pressure senses. If, as seems reasonable, one assumes that these senses can provide (for example, in the course of acting upon objects) input that is interpretable as information about FORCE, then the infant has a further valuable source of evidence about the physical world. Notice that this proposal endows the infant with the perception of mechanical properties of objects and events - things objectively in the world - rather than merely with a subjective awareness of bodily sensations. This general idea receives support from the findings of Streri and Spelke (1988) on the haptic perception of object cohesion.

A second line of investigation suggested by **ToBy** concerns infants' knowledge about types or *kinds* of objects. According to the theory of **ToBy**, perceptual categorization of objects is distinct from other forms of knowledge about object-kind. The three-dimensional model catalogue, for example, could provide the basis for perceptual categorization of purely visual objects. But what about an infant's knowledge of object-kinds as it relates to the mechanical object? The results on understanding hands as sources of power suggests that this is one object-kind that infants do known about and distinguish from other kinds of objects. Little other than this is known at present. The theory of **ToBy** predicts that mechanical properties will provide the central information in the early formation of (conceptual) object-kinds. As we shall see in the following section, information about how Agents typically use an object will come to assume importance too. Acquiring this information demands more than mechanical FORCE descriptions of Agents. A new level in the theory of Agency is required and by assumption a further mechanism beyond **ToBy.**

Beyond ToBy

I come now to the limitations on **ToBy** as a theory of Agency. The mechanical properties of Agents are not the only properties that set them apart from other objects. When Agents behave they act in pursuit of goals and react to their distant environment (perception). Neither of these facts fits in with or is representable by **ToBy's** contact mechanics. A goal is a state of affairs that an Agent tries to bring about. Typically, this state of affairs does not yet exist - the Agent acts to make it actual; the desired state of affairs is a little in the future, one might say. Sometimes action in pursuit of a goal fails and the strived for state of affairs does not come about - it remains "in the future." Understanding the pursuit of a goal by an Agent, then, is understanding action in relation to circumstances that are at a distance in time. Understanding perception, on the other hand, typically involves understanding a causal relation at a distance in space. Neither of these fits with **ToBy's** spatiotemporal contact principle of "no action at a distance." **ToBy** is concerned with mechanical relations that obtain locally and contiguously in space and time. The brain mechanisms implementing **ToBy,** then, will presumably operate at this strictly local scale. Understanding Agents and Action,[9] however, requires an analysis over a larger scale to capture relations between Agents and states of affairs at "distant" times and places. I assume, therefore, that actional (as opposed to mechanical) properties of Agents are not processed by **ToBy** but by a different mechanism, operating after **ToBy.** I call this mechanism **ToMM. ToMM** is concerned with the *intentional* properties of Agents.

Agency and the "fictional causes" problem

We come now to what seems to me to be the really difficult problem that the existence of Agents puts in the way of an information processing system that wants (blindly of course) to evolve the capacity to understand their behavior. Up to now, we have considered only the mechanical domain. When considering merely physical objects, only *actual* circumstances are relevant in accounting for *actual* behavior; only really present objects can possibly be relevant to prediction; only states of affairs that actually obtain can possibly contain actual causes of actual behavior. This is perhaps the most basic and truistic assumption of all in causal domains: that only real causes real. Without this assumption, there is no notion of cause, there are no constraints on possibility, no causal explanation, no causal prediction, in short, no causal knowledge. To suppose that something unreal caused something real does not even begin to make sense. And yet, here is an example of a kind of reasoning we do everyday: Why did John jump into the doorway? To avoid the rain. Was it really raining? No, he only thought it was raining.

Because Agents have cognitive properties (as well as mechanical and actional properties), they sometimes (often) behave in response to situations that are not actual but are merely fictional. This sounds like we assume that something fictional is causing something actual. Since that is crazy, I am going to assume that that is not what we are doing. The problem that **ToMM** solves is how to describe the relation that holds between the actual behavior of Agents and fictional circumstances while maintaining a causal, that is, rational framework. **ToMM**'s job is to square this circle.

A short digression on actional Agency

In describing the actional properties of Agents above, I touched on a weak version of the fictional causes problem in the notion of goal-directed action. In this case, however, there is no reason to see the goal state of affairs as causing the behavior. Instead, just as in mechanical Agency, where the Agent is seen as having an internal source of causal power or FORCE, so in actional Agency is the Agent seen as a having an internal source of striving or acting (toward the goal). The action or "pressure" to bring about the goal state of affairs arises from *within* the Agent and "flows outward" from the Agent (see Talmy, 1988, for a discussion of how basic verb structures in English express this kind of picture).

The notion of the Agent acting or striving to bring about a goal state is akin to Wellman's (1990) attribution of a "drive theory" of desire to the young child. Wellman is right in postulating an early notion of desire that is not propositional attitude based. However, the way he formulates his drive theory cannot be correct. Wellman says the child simply conceives of another person as having an internal drive toward an object but without the possibility of embedding this object in a proposition, as it were. Wellman depicts the child as imagining a pair of hands in the other person's head stretched out to an apple, "wanting" the apple. In denying the child the ability to represent any proposition-like content as the focus of the desire state, Wellman wishes to avoid attributing a propositional attitude notion of desire to the very young child. However, his formulation has the unfortunate effect of making the child's putative drive notion almost useless for predicting behavior. When Wellman's child thinks of Mary as wanting an apple, he is incapable of representing what Mary wants to do with the apple or what state she wants the apple to be in. Somehow Mary just "wants" an apple. Of course, the reader of Wellman's theory will involuntarily supply an enriched interpretation to Mary's "want": by default, that Mary wants to *eat* the apple. But the young child, according to Wellman, cannot do this. When Wellman's child thinks that Billy wants a swing, he cannot represent whether Billy wants to sit on the swing, swing on the swing, just sidle up and be close to the swing, or anything else specific. To do so would be to represent Billy as desiring a state of affairs rather than just an object. Billy simply "wants" the swing, full stop. Unfortunately, such a notion is pretty useless for predicting behavior.

However, there is another way to formulate Wellman's insight that there may be a notion of desire that is not based on a propositional attitude, but one gets to this not by dropping the state of affairs (described by the proposition in a propositional attitude), but by dropping the attitude. In its simplest form this notion represents the Agent as "ACTING to bring about [a state of affairs]." In a more complex form this notion might represent the Agent as having a disposition or standing readiness to ACT to bring about [a state of affairs]. Further elaborations of this general class of notion include disposition strength, preference, ACTING to avoid [a state of affairs], and so on. The state of affairs in brackets would, of course, be represented by the child however he normally represents states of affairs, and if this is normally represented in a "propositional" format then so be it. Such a notion, however, is not a propositional attitude because "ACTING to bring about..." does not describe an attitude to the truth of a proposition; it describes an attempt to change physical circumstances. It will be necessary for the child to be able to represent possible or future states of affairs, but then this is already assumed by the kind of understanding of the *physical* world that has been demonstrated in babies, as discussed in earlier sections.

Attitudes and fictions

In actional notions of Agency, one sees only a weak form of the fictional causes problem. In the case of propositional attitudes like *believing* (and *pretending),* one sees the full-blown fictional causes problem. Although the theory of Agents and Action allows one to understand behavior in relation to circumstances that are remote in time and space, it does not allow one to construe behavior in relation to (caused by) circumstances that are fictional *in the here and now.* Interpreting behavior in this way is made possible by the third and final layer of the theory of Agency: Agents and Attitudes. The notion of a propositional attitude solves the fictional causes problem. It does so essentially by describing the Agent as actively holding an attitude to the truth of a proposition.

In a successful mental state attribution, the state of affairs described by the proposition that the Agent is related to may be actual, merely possible, or even impossible. But the attitude the Agent takes (to its truth) is always real.

For example, John *believed* it was raining. The rain was not real but John's attitude of believing in the rain certainly was real. And it is John's attitude (to the truth of the proposition in question) that caused his behavior (not the proposition nor the state of affairs described by the proposition).

There is, of course, a price to be paid for solving the problem posed by the fact that the behavior of Agents is determined by cognitive properties.[10] The price to be paid is that special concepts have to be available − that is to say, a specialized representational system is required. I have called this system the "metarepresentation" (Leslie, 1987b) or, more recently, the "M-representation" (Leslie & Thaiss, 1992; Leslie & Roth, 1993). In the closing pages, I briefly sketch how **ToMM** might come to solve these problems.

To summarize: **ToMM** itself has an internal structure with at least two major subcomponents, which I shall call *system*$_1$ and *system*$_2$. *System*$_1$ deals with the second level of the theory of Agents, namely, Agents and Actions. *System*$_1$ employs the "ACT to bring about state of affairs" system of representation. *System*$_2$ implements the third level of the theory of Agents, namely, Agents and Attitudes. *System*$_2$ employs the M-representation.[11]

System$_1$*: Agents and Action*

Whereas **ToBy** appears to begin development at around 3 to 4 months, the **ToMM** *system*$_1$ probably begins to develop later, perhaps around 6 to 8 months.[12] One of the first signs of *system*$_1$ is the following of eye gaze. From about 6 months onward, infants will begin to turn and visually search in the direction of gaze adopted by an *en face* adult (Butterworth, 1991). Butterworth found that the accuracy of the infant in locating the object that the adult is looking at increases rapidly between 6 and 12 months. Also during this time, infants begin to attend to the uses Agents make of objects; by 12 months infants display knowledge of the conventional function of some everyday objects, such as spoons and brushes (Abravanel & Gingold, 1985). This suggests that they are able to appreciate and are storing knowledge about the instrumental roles that objects can play in Agent's goal-directed actions.

If *system*$_1$ underlies the infant's ability to represent the actional properties of Agents, then it really comes into its own when representing situations in which two Agents come together and *interact.* Agent's goals can mesh in different ways. At a general level: One Agent's goal may coincide with another Agent's goal. This produces enhancement or helping. Alternatively, one goal may be opposed to another. This produces blocking or harming (cf. Premack, 1990). More specifically, during the second half of the first year, infants begin to produce "requesting" behaviors, for example, in request reaching or "asking" to be picked up (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Bruner, 1976). These requests call upon another person to achieve or help in achieving some current goal of the infant's. Similarly, during this time, infants begin deliberately to acquiesce in or comply with the actions of others, for instance, positioning to be picked up. Also during this time, they begin deliberately to try to block the goals of the other person, such as in "refusing" behaviors (Bates et al., 1979; Bruner, 1976). Again during this time, infants begin to show they appreciate the role structures of some simple goal-directed (inter)actions by being able to reverse the roles, for example, in the "give and take" behaviors that appear around 10 months (Bruner, 1976).

Premack (1990) makes a number of interesting suggestions concerning infants' sensitivity to goal-directedness. One suggestion is that goal-directedness or "intentionality" may be perceived on the basis of a class of motion pattern. If so, this would parallel the case of mechanical causality, for example, launching. In terms of the present framework, Premack suggests a specialized device that renders a "purely visual goal-directedness." This is an interesting idea that, if true, would comport with and extend the examples discussed earlier.

It remains very much an open question whether infants bring some simple theory of the kinds of goals Agents may have or whether all this must be learned. The simple representational system I have proposed for *system*$_1$ will at least allow the infant to learn about some immediate goals Agents may have by watching for outcomes. The outcome state of affairs can then be entered into the action representation as the goal state of affairs. Even in this limited scheme, outcome information can be useful, given some experience, for construing later actions of Agents that are directed to the same kind of goal, even when the intended outcome is not achieved.

System$_2$*: Agents and Attitudes*

ToMM *system*$_2$ begins to develop last, during the second year of life. Perhaps the clearest early sign of the employment of the M-representation is the ability to pretend and understand pretense in others (Leslie, 1987). This emerges between 18 and 24 months. During this period, infants become able to construe the behavior of other Agents as relating to fictional states of affairs, specifically, as issuing from the attitude of pretending the truth of a proposition that describes a fictional state of affairs. For example, a mother's actual behavior of talking to a banana can be understood by constructing the M-representation, **mother pretends** (of) **the banana** (that it is true that) **"it is a telephone."** This links her behavior, via an attitude, to a fiction. For fuller discussion see Leslie (1987), Leslie and Thaiss (1992) and Leslie and Roth (in press).

Shared pretense can be properly regarded as an example of intentional communication (Leslie & Happé, 1989). What is particularly interesting about this communication is that it forces the child to compute *speaker's meaning* in Grice's (1957) sense. Thus, mother continues pretending and says, "The telephone is ringing. It's for you." She hands the child the banana. To comprehend this otherwise bizarre behavior, the child must not construe mother's words simply in terms of their linguistic meaning. She must also compute what mother intends them to mean on this occasion. In this example, "telephone" is intended to refer to the banana whereas "is ringing" is intended as a predicate true not directly of the remarkably silent banana but only of the banana via its pretend identity as a telephone. If the child does not compute speaker's meaning in relation to speaker's mental state (i.e., speaker pretends that...), she will fail to comprehend mother's actions, linguistic and otherwise.

Before the emergence of pretense, however, from around 12 months onward, the infant already communicates informatively with other Agents and understands *some* of the informative communications other Agents direct at him or her. These new communications go beyond the "instrumental" request and other earlier instrumental communicative behaviors of the first year that are under the control of $system_1$ (and possibly, in certain cases, of $system_0$ [see Note 11]). For example, the infant begins to understand informative pointing gestures around 14 months (Blake, McConnell, Horton, & Benson, 1992; Butterworth, 1991). Informative *showing* also typically makes its appearance early in the second year along with verbal communication.

There are a number of quite tricky questions that this brisk summary skips over. For example, what is the relation between linguistic abilities and **ToMM**? My assumption is that structural linguistic knowledge and language processing mechanisms are essentially independent of **ToMM. ToMM's** development may impact, however, on communicative language use.

Another set of questions concerns the development of attention to Agents, joint attention with Agents and gestures, such as pointing and informative showing, directed to Agents. In one guise, **ToBy** and **ToMM's** sub-systems form a hierarchy of control mechanisms governing attention and responsiveness to Agents. The critical point in understanding the development of attention to Agents, joint attention, and social responsiveness over the first two years is how these hierarchical control mechanisms develop. A given response may have different significance depending upon its controlling mechanism. For example, though we begin to smile after the first few weeks of life and continue smiling (on and off) for the rest of our lives, *what* we are smiling *at* changes drastically with development. The 1-week-old smiles only when asleep, the awake 6-week-old smiles at high-contrast stimuli, months later the infant will direct smiles at familiar animated faces but not at other high-contrast stimuli, and still later when sharing pretense, and so on. The development of "the smile response" is really the development of its controlling mechanisms. Similar considerations no doubt hold for other social responses such as pointing, showing, vocalizing, joint attention, and so forth. One way to study these questions is to examine those tragic cases where brain growth proceeds abnormally as in childhood autism. Some of these issues are explored at greater length in Leslie and Roth (1993).

In general, the developments in relation to messages and communication discussed in this section are indicative of the 1-year-old's new approach to Agents as possessors, transmitters, and recipients of *information.* Viewing Agents as transmitters of information, and not just FORCE, heralds the beginnings of the capacity to solve the problems created by the fact that in the real world Agents' behavior is determined by cognitive properties. The solution to this problem hinges on understanding how *meaning* enters into the causation of behavior. The social intelligence that now dominates this planet is the result of the evolution of neural mechanisms that rapidly find this solution.

Summary

I have presented a three-level theory of our understanding of Agency. Each of these levels corresponds to an information processing subsystem specialized for making certain kinds of information explicit. Each subsystem attends to one of the three classes of properties that distinguish Agents from non-Agents. These classes are: The behavior of Agents reflects the mechanical property of having an internal source of energy; the behavior of Agents reflects the actional property of pursuing goals and perceiving the environment; and finally, the behavior of Agents is determined by cognitive properties. It seems likely, given the distinctive computational demands of tracking the world at each of these levels, that correspondingly distinct devices are required in cognitive architecture. Agency as a domain of knowledge, then, has a complex structure. Understanding conceptual development in this domain requires understanding how this structure reflects core architecture − those aspects of the organization of human information processing that form the basis for development rather than its outcome.

Notes

1. I hope the reader will simply sound out these acronyms like a name rather than internally spell out what they stand for every time they are read. This way they will come to sound like old friends.
2. One possibility is that ToBy's initial notion of object is parallel to the notion of an "object file" (Kahneman & Treisman, 1984), that is, a spatiotemporally continuous entity none of whose properties, such as size or shape, are considered critical, or even relevant, for determining its identity. The "file" itself is what allows the object to be referred to. The various properties that might be associated with the object are gathered together and "held" in the file but are regarded as accidental to the identity of the object. The idea is like "rigid designation" (Kripke, 1972): referring to *this* thing whatever it is. Such a parallel would have to be qualified so that **ToBy**'s object files applied only to three-dimensional

objects. However, ToBy may employ some initial idea of object type and/or be ready to assign objects to kinds. In this case, ToBy might try to decide questions of identity with respect to kind. In the case of object files the question is, Is this the same individual thing (whatever it may be)? versus, in the case of object kinds, Is this the same one of a given kind? I leave this as an open empirical question about which we know very little from studying infants, though I will suggest below that there is at least one object kind infants know about early on (viz., hands).

3. Schlottmann and Shanks (1992) provide further evidence for a Michotte module in adults. They forced their subjects to attend to the predictive relationship between the movements of the objects in collision events. The contingency of the second movement on the first was varied and subjects' predictive judgments showed they were sensitive to the degree of contingency. However, contingency had no effect whatsoever on their "causal perception" of single launching events. The illusion of a mechanical cause was highly sensitive to *contiguity* whereas predictive judgments were much less so. Schlottmann and Shanks take the dissociation between contiguity and contingency as support for "a distinct mechanism of causal perception" (p. 340). In terms of ToBy, what determines judgments of mechanical cause in collision events is how well the conditions for the trans mission of FORCE have been met, not the statistics of the event set. Use of statistical relations in judgments of cause probably assumes importance to the extent that the analysis of mechanism becomes more difficult, less visible, and less complete.
4. So far as I can see, nothing of what I say hinges on the details of Marr's "three-dimensional model" account of object recognition; other accounts in the same general class, such as Biederman's (1987) theory of "geons," are equally compatible.
5. I use the term *spatiotemporal* in preference to Mandler's use of the term *spatial* since this better expresses the importance of motion in her account.
6. One might gain the impression from reading Mandler (1992) that image schemas abstract over shape and volume such that what is made explicit is a two-dimensional spatial arrangement, so that containment is represented as an *x* inside a circle. However, Mandler's intention is that the schema be a kind of three-dimensional model of the entity and container. In either case, whatever spatial properties of the INSIDE relation an image schema captures, it is only in the case of three-dimensional objects that INSIDE becomes mechanical containment. That is, when three-dimensional objects are involved, the spatial aspect of containment can take on mechanical significance and be consequential: The contained object cannot escape by passing through the walls of the container; thus when the container is moved it will transport the contained object to which ever location the container is moved; the object can however enter and escape through an aperture in the container, if certain metrical conditions are met, and so forth. These mechanical consequences follow from (1) the spatial relation, INSIDE, *together with* (2) certain assumptions drawn from the theory of object contact mechanics, such as solidity and transmission of FORCE (in this case, entrainment or transport). Infants around 6 months, perhaps earlier, already appreciate such mechanical containment (Leslie & DasGupta, 1991). But in the spatial image schema these mechanical properties are not made explicit, in Marr's

sense. Even if the image schema does depict a three-dimensional relationship over time, it still fails to make the mechanical relationships explicit. To achieve this, mechanical properties, such as solidity and transmission of FORCE, would have to be added on to the image. The image schema is therefore unlikely to be the primary basis for conceptual development.

7. Machines do not pose much of a problem for **ToBy.** However, given that they could not have been taken into account during evolution, they might well have been a problem. For example, a robot will be considered an Agent under this scheme, though we would not consider it animate. My assumption, you will recall, is that **ToBy** knows nothing of animateness. Automobiles might appear to be Agents to the extent that they keep changing motion state (but the first time an infant saw a car simply rush past, she might assume that something external has propelled it, like a giant billiard ball). On the other hand, an indistinct blob rushing about, such as a housefly, will usually (and correctly) appear to be an Agent without necessarily appearing to be animate. Separating Agency from animateness (by evolving a modular organization) has had beneficial side-effects, allowing us to apply without obstacle our commonsense understanding of Agency (1) to inanimates and (2) without having to know whether or not something is animate.

   Picking out an object as an Agent has a cascade of effects. As we shall see later in this chapter, to the extent we think something is an Agent, we are willing to try to attribute goals to its behavior, and to the extent we are able to attribute goals to the Agent, we are willing to attribute propositional attitudes to explain, predict, and interpret that behavior. For example, houseflies try to get out through windows because they don't *know* there's glass in the way.

8. Vaina (1983), working within a similar framework to that adopted here, outlines a theory of the "Functional" representation. This deals with the representation of the functional properties of objects such as their "throwability," "pick-upableness," and so on. Although Vaina is careful in her consideration of the kinds of visual information made explicit prior to and forming the input to the computation of the Functional representation, in my view her theory overlooks or gives insufficient weight to two critical things. The first is the necessity to make explicit the mechanical properties of objects and scenarios. The second is to make explicit the actional properties of Agents (see the discussion in the following sections). The functional properties of objects are determined in part by their mechanical properties and beyond this by the kinds of uses Agents make of them in pursuing goals.

9. This makes a nice phrase but I intend to subsume perception under it too. Actually, only perception in a limited sense is subsumed. For example, the relation of "seeing x" is subsumed but not the relation of "seeing that p." The latter falls under Agent and Attitude.

10. Strictly speaking, not *all* the behavior of an Agent is determined by cognitive properties. Mechanical behavior, like being run down by a bus, is not so determined. The *actions* of Agents are determined by cognitive properties.

11. To avoid misunderstanding I should say that neither **ToMM** alone nor **ToMM** and **ToBy** together exhaust social intelligence. For example, mechanisms of face recognition lie outside the systems discussed here but clearly form part of the capacity for social responsiveness. It seems highly likely that such low-level

mechanisms have inputs into **ToMM.** For this reason one might lump together such mechanisms under the heading of $system_0$. There are many interesting possibilities in this regard but I shall not pursue them here.

12. Although the particular ages I mention are not critical for the theory, they seem to me as a matter of fact to be approximately right.

References

Abravanel, E., & Gingold, H. (1985). Learning via observation during the second year of life. *Developmental Psychology, 218,* 614-623.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6- and 8-month-old infants. *Cognition, 23,* 21-41.

Baillargeon, R. (1991a). Infant's understanding of the physical world. Paper presented to symposium, "Infants' reasoning about spatial relationships," SRCD Biennial Conference, Seattle, April 1991.

Baillargeon, R. (1991b). The object concept revisited: New directions in the investigation of infants' physical knowledge. In H. W. Reese (Ed.), *Advances in Child Development and Behavior, Vol. 23.* New York: Academic Press.

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

Bates, E., Benigni, L., Bretherton, I., Camaioni, L, & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York: Academic Press.

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

Blake, J., McConnell, S., Horton, G., & Benson, N. (1992). The gestural repertoire and its evolution over the second year. *Early Development and Parenting, 1,* 127-136.

Bruner, J. S. (1976). From communication to language - a psychological perspective. *Cognition, 3,* 255-287.

Butterworth, G. (1991). The ontogeny and phylogeny of joint visual attention. In A. Whitten (Ed.), *Natural theories of mind: Evolution, development, and simulation of everyday mind reading.* Oxford: Basil Blackwell.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition, 41,* 83-121.

Chomsky, N. A. (1975). *Reflections on language.* New York: Random House.

Chomsky, N. A. (1986). *Knowledge of language: Its nature, origin, and use.* New York: Praeger.

Dennett, D. C, & Haugeland, J. C. (1987). Intentionality. In R L. Gregory (Ed), *The Oxford companion to the mind* (pp. 383-386). Oxford: Oxford University Press.

Diamond, A. (1988). Differences between adult and infant cognition: Is the crucial variable presence or absence of language? In L. Weiskrantz (Ed.), *Thought without language* (pp. 335-370). Oxford: Oxford Science Publications.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107-117.

Ellis, H. D., & Young, A. W. (1989). Are faces special? In A. W. Young and H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 1-26). Amsterdam: Elsevier Science Publishers.

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

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction. *Cognitive Science, 14,* 79-106.

Grice, H. P. (1957). Meaning. *Philosophical Review, 66,* 377-388.

Hume, D. (1740). *A treatise of human nature.* London: Clarendon Press.

Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development.* Oxford: Blackwell.

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

Kripke, S. A. (1972). Naming and necessity. In D. Davidson and G. Harman (Eds.), *Semantics of natural language* (253-355). Dordrecht: Reidel.

Leslie, A.M. (1982). The perception of causality in infants. *Perception, 11,* 173-186.

Leslie, A. M. (1984a). Infant perception of a manual pick-up event. *British Journal of Developmental Psychology, 2,* 19-32.

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

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review, 94,* 412-426.

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. (unpublished). Further observations on the Pulfrich double pendulum illusion. MRC Cognitive Development Unit, University of London.

Leslie, A. M., & DasGupta, P. (1991). Infants' understanding of a hidden mechanism: Invisible displacement. Paper presented at symposium on "Infants' reasoning about spatial relationships," SRCD Biennial Conference, Seattle, April 1991.

Leslie, A. M., & Happé, F. (1989). Autism and ostensive communication: The relevance of metarepresentation. *Development and Psychopathology, 1,* 205-212.

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

Leslie, A. M., & Roth, D. (1993). What autism teaches us about metarepresentation. In S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 83-111). New York: Oxford University Press.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43,* 225-251.

Livingstone, M. S., & Hubel, D. H. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240,* 740-749.

Mandler, J. M. (1992). How to build a baby, II: Conceptual primitives. *Psychological Review, 99,* 587-604.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman & Co.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial

organization of three-dimensional shapes. *Proceedings of the Royal Society of London, B, 200,* 187-217.
Michotte, A. (1963). *The perception of causality.* Andover: Methuen.
Michotte, A., & Thines, G. (1963). La causalite perceptive. *Journal de Psychologie Normale et Pathologique, 60,* 9-36. Translated by G. Thines, A. Costall, and G. Butterworth in G. Thines, A. Costall, and G. Butterworth (Eds.), *Michotte's experimental phenomenology of perception* (pp. 66-87). Hillsdale, NJ: Erlbaum. (1991).
Piaget, J. (1953). *The origins of intelligence in the child.* London: Routledge and Kegan Paul.
Piaget, J. (1955). *The child's construction of reality.* London: Routledge and Kegan Paul.
Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure.* Cambridge, MA: MIT Press.
Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707-784.
Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36,* 1-16.
Schlottman, A., & Shanks, D. R. (1992). Evidence for a distinction between judged and perceived causality. *Quarterly Journal of Experimental Psychology, 44,* 321-342.
Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47,* No. 1.
Spelke, E. (1988). Where perceiving ends and thinking begins: The apprehension of objects in infancy. In A. Yonas (Ed.), *Perceptual development in infancy. Minnesota Symposium on Child Psychology* (pp. 191-234). Vol. 20. Hillsdale, NJ: Erlbaum.
Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14,* 29-56. Streri, A., & Spelke, E. (1988). Haptic perception of objects in infancy. *Cognitive Psychology, 20,* 1-23.
Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12,* 49-100.
Tanaka, J. W., & Farah, M. J. (1991). Second-order relational properties and the inversion effect: Testing a theory of face perception. *Perception & Psychophysics, 50,* 367-372.
Treisman, A. (1988). Features and objects: The fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 40,* 201-237.
Vaina, L. (1983). From shapes and movements to objects and actions. *Synthese, 54,* 3-36.
Van Essen, D. C. (1985). Functional organization of primate visual cortex. In A. Peters and E. G. Jones (Eds.), *Cerebral cortex* (Vol. 3) (pp. 259-329). New York: Plenum.
Warrington, E. K., & Taylor, A. M. (1978). Two categorical stages of object recognition. *Perception, 7,* 695-705.
Warrington, E. K., & James, M. (1986). Visual object recognition in patients with right-hemisphere lesions: Axes or features? *Perception, 15,* 355-366.
Wellman, H. M. (1990). *The child's theory of mind.* Cambridge, MA: MIT Press.