RuCCS TR−12
November, 1993

A Theory of Agency

Alan M. Leslie

Technical Reports
of the
Rutgers University
Center for Cognitive Science

A Theory of Agency

Alan M. Leslie
MRC Cognitive Development Unit
University of London
and
Center for Cognitive Science
Rutgers University

A Theory of Agency

Alan M. Leslie

I shall outline a “tri-partite” theory of our core understanding of Agency. More specifically, I shall discuss a theory of the core constraints that organise our early learning about the behaviour of Agents.

To clear the ground a little, I should mention a couple of things I do *not* mean by “Agency”. Sometimes, “agency” means just any cause of an event. But what I mean by Agency is a particular class of *object*. Often the term “agent” is used to denote the entity that is momentarily the cause of some other event—e.g. the stone that breaks the window. This is *not* how I shall the term. Instead, an Agent (with a capital letter) is a type of object and Agency an enduring property of that object. Clearly, people are examples of things which are enduringly Agents, even though they may, from time to time, be the passive victims of some particular event. In this sense, Agency is a property of objects rather than a property of motion.

The second thing I do not mean by Agency is animacy. The type of model I will develop appeals to a notion of modular design (Marr, 1982). In the context of core commonsense conceptual knowledge, modular design translates into the notion of “sub-theory”. Core knowledge is parcelled into subtheories to the extent that an information processing system is organised into modules. From this point of view, the concept, *animate being*, carries biological connotations—is living, reproduces, grows, etc.—that is, by hypothesis, extrinsic to our core notion of Agency. The system I describe does not make explicit information of a biological character. I leave it as an open empirical question whether some other part of core systems, not explored here, represents biological information (see chapters by Keil and by Carey, this volume).

The central part of the theory of cognitive development deals with core cognitive architecture—it characterises those properties of the information processing system that provide the basis for development, as opposed to those properties that are the result of development. For example, in classical associationism core cognition is assumed to consist purely of statistical associative processing over elementary “sensations”. In other views, core architecture is assumed to have a more varied componential character. Core structure will, in the main, reflect specialization for carrying out particular information processing tasks as a result of adaptive evolution. From this point of view, Agents have provided a potent source of adaptive pressure in human evolution with obvious ‘benefits' in terms of enhanced social intelligence flowing from a sophisticated capacity to explain, predict and interpret their behaviour.

The tripartite theory of Agency begins with the idea that Agents are a class of objects possessing sets of causal properties that distinguish them from other physical objects. My next assumption is that, as a result of evolution, we have become adapted to track these sets of properties and to efficiently learn to interpret the behaviour of these objects in specific ways. My story begins by identifying three classes of distinctive properties of Agents that determine Agents' behaviour. Each of these classes of properties sets Agents apart from mere physical objects. And each of these classes of properties produces its own special problems for a processing system whose task is to understand Agents' behaviour.

Three sub-theories of Agency

Three classes of real world properties distinguish Agents from other physical objects. These are:

1) *Mechanical properties*: Agents have mechanical properties that mere physical objects do not have. I shall characterise the main difference in terms of having an internal and renewable source of "energy" or FORCE, versus not possessing such a source and thus having to rely on external sources.

2) *Actional properties*: Agents do not simply move and take part in events. Agents *act* in pursuit of goals and re-*act* to the environment as a result of perceiving. Mere physical objects do not act in pursuit of goals and do not perceive their environment. Furthermore, the acting and re-acting Agent can come together with another Agent and inter-*act*.

3) *Cognitive properties*: The behaviour of Agents is determined by cognitive properties, e.g., holding a certain attitude to the truth of a proposition. Mere physical objects do not have cognitive properties.

My claim is that tracking each of these three classes of properties poses to some extent distinct information processing problems that require to some extent distinct solutions. My hypothesis is that, as the result of the evolution of a modular design, our core notions of Agency reflect three distinct processing mechanisms arranged hierarchically. Succeeding mechanisms interpret Agents' behaviour at succeeding levels of representation. Each level corresponds to a different "sub-theory" of Agency. Description at one level provides the principal relevant input for inferring the appropriate description at the next level. These three mechanisms with their respective "sub-theories" introduce in turn the three causal *paradigms* that form the core of human commonsense: "mechanical causality", "teleological causality", and "psychological causality".

Mechanics and the FORCE Representation

David Marr pointed out that any representational system makes explicit certain kinds of information at the expense of backgrounding other kinds of information. Marr illustrated this idea by comparing Roman and Arabic numeral systems in terms of the extent to which they make arithmetic relations explicit. The key question to ask about a representational system is what sorts of information it makes explicit in contrast to the other kinds of information it leaves implicit. For example, earlier I suggested that our core representations of Agency left it implicit whether Agents possessed biological properties. In this section, I will argue, contrary to the traditional assumption which reduces all physical knowledge to spatiotemporal information, that there is a level of representation that is basic to human intelligence that makes mechanical information explicit.

David Hume (1740) was to a large extent responsible for the diversion of attention from commonsense mechanics that has characterised much of modern psychological thought. Hume's programme of doing away with a core level of mechanical understanding was to become a central plank of the empiricist programme. Core understanding, according to this view, comprised essentially two things. First the registration of spatio-temporal properties; and secondly the statistical association of spatio-temporal properties with one another. Mechanics was squeezed out of the system. Hume begins his analysis by considering the launching event.

> "Here is a billiard ball lying on the table, and another ball moving toward it with rapidity. They strike, and the ball which was formally at rest now acquires a motion. This is as perfect an instance of the relation of a cause and effect as any which we know either by sensation or reflection." (Hume, 1740/1962:292)

Hume goes on to argue that the only thing we see in such an event is its spatio-temporal properties: namely, the spatial contact between the two objects and the movement of the second object without delay after impact by the first. We can not, Hume argued, see, that is, *observe*, the causal relation. This meant to Hume that our idea of causal relation had to be the result of statistical associations formed within this recurring spatiotemporal pattern. I believe that Hume was right about the first part of his analysis, namely that we *see* only the spatiotemporal properties in this event. He was right about seeing only spatiotemporal properties—not because his theory of *observation* was right, but simply because that is the job of *vision*. Vision makes explicit information about space and spatial arrangement over time. This is what Marr (1982) called the "quintessential fact about vision". But Hume was wrong to conclude that our idea of causal relation must therefore be based on statistical association. As we shall see, what makes launching seem "perfect" to us as an instance of cause and effect, is that it instantiates a mechanical interaction with a perfect transmission of "FORCE."

Launching and FORCE

FORCE is a primitive mechanical notion, invoked by attention to 3-D bodies and their arrangements. It must be stressed that FORCE is not be to identified with the scientific notion of energy or force or any other scientific notion. The FORCE representation is the result of evolutionary adaptation and not the product of a scientific theory developed culturally. The general idea behind the FORCE representation is (a) when objects move, they possess or bear FORCE and (b) when objects contact other objects, they transmit, receive or resist FORCE. Talmy (1988) has discussed a similar notion of "force dynamics" discernable in patterns of lexicalisation and verb argument structure in natural language.

The FORCE representation has an important role in determining the sorts of things infants know or assume about the physical world. Let me begin by looking at three phenomena from my own work on infancy which support and illustrate this idea. I shall then point to some key findings from the research of Spelke and of Baillargeon that also support the assumption of a core FORCE representation. In a series of experiments some years ago (Leslie, 1982, 1984b; Leslie & Keeble, 1987), I showed that six-month-old infants were sensitive to the causal structure of the kind of billiard ball launching effect David Hume had described.

Figure 1 about here

In the critical experiment (Leslie & Keeble, 1987), infants were habituated to a film of a direct launching event (see Figure 1). A second group of infants was habituated to a variation on this launching event in which a short time delay is interposed between the impact of the first object and the movement of the second object. Michotte (1963) had shown that for adults the interposition of this short delay destroyed the impression of causality or pushing that was clear when viewing the direct, non-delayed event. When both groups of infants had been habituated to their respective events, they were each tested on exactly the same event respectively, but played backwards in time.

From a *spatiotemporal* point of view, both groups of infants on testing saw equal changes. For both groups of infants, their test event now occurred in an opposite spatial direction. Also for both groups, the event now played with an opposite temporal order. Likewise for both groups, whatever degree of motion continuity present during familiarisation was present once again during the test. Finally, the contingency between the motions in the familiarisation events was present again in the test events. All this stems from the fact that both groups of infants simply saw the same film they had seen previously but with the projector running in reverse.

From the point of view of *mechanics*, however, there was an important difference between the effect of reversal on the two initial films. In the causal, direct launching event, the two objects appear to have different mechanical roles: the first object is a transmitter of FORCE, a pusher, while the other is the recipient of FORCE, an object which is pushed. When this event is reversed these mechanical roles are also reversed: the pusher becomes the pushed. In the case of the non-causal event, however, there is no such apparent mechanical relation between the objects, there is no pusher and no pushed. Thus, when this event is reversed there is still no pusher and no pushed and consequently no reversal of roles. If the infants construe these events from a mechanical point of view then the direct launching event will be more interesting in reverse than the non-causal delayed event in reverse. So even though the spatiotemporal changes and the contingency properties are equated in the test for the two groups, the causal group should recover attention more. This is exactly what we found.[1]

The above results are immediately understandable in terms of a mechanical or FORCE representation which infants tacitly deploy. Furthermore, a FORCE description allows us to understand, where Hume could not, why launching should seem "perfect" to us as an instance of cause and effect. Consider how the FORCE representation describes the behaviour of the objects in the launching effect. The event begins with object A moving and object B stationary. Object A therefore has to be the initial bearer of FORCE rather than object B. When object A impacts on Object B and becomes stationary, it transmits all its FORCE to Object B. Object B which is now in motion has gained the FORCE formerly borne by Object A.

The above findings with six-month-old infants depend upon the infants seeing a particular causal, presumably mechanical, *direction* in the launching event. Our results showed that the causal direction is conceptually distinct from the spatiotemporal direction. However, having a definite causal direction is not the inevitable result of just any kind of mechanical construal. In fact, there is a causal symmetry[2] if the event is construed in terms of Newtonian mechanics: when object A impacts on object B and causes B to move off there is an opposite and equal reaction from B to A. Under reversal the same symmetrical relationship would hold again.

---

[1] Cohen (Cohen & Oakes, 1993; Oakes & Cohen, 1990) has criticised the claim that these results demonstrate a "modular perception" in infants. Cohen seems to me to misconstrue some aspects of the problem. For example, Cohen obtained evidence of a causal perception in 10-month-olds but not in 6-month-olds. The striking finding, however, is that whenever the infants in his study showed sensitivity to the spatiotemporal properties of the particular stimuli he used, they also showed sensitivity to causal properties. Having said that, the particular version of the "modular theory" that Cohen attacks (e.g. Leslie, 1988) is not the same in important respects as that argued for here. It may be that some of Cohen's recent results fit better with the present "theory of body" view.

[2] I am indebted to Liz Spelke for this point.

However, commonsense intuition is that being stationary and moving at a constant velocity are fundamentally different states. The results on the perception of launching suggest that infants share this intuition. We cannot tell for sure from these results which direction the infants apprehend the causal direction running in: whether it is B's moving away from A's impact that causes A to stop, or, what seems much more likely, that A's impact causes B to move away. I put forward the following hypothesis. Thew infant's "theory of body" (**ToBy**) tacitly employs the idea that FORCE is transmitted from one object to another in a particular direction, from the moving to the stationary, like a baton being passed on in a relay race. There is no opposite and equal reaction as far as **ToBy** is concerned. This defines *mechanical* direction, the direction of FORCE transmission. Mechanical directionality in turn gives rise to the asymmetrical notions of mechanical role that seem so ubiquitous in commonsense. Our sensitivity to launching, both as adults and as infants, is in part structured around the notion of mechanical role, which, in turn, reflects the direction of FORCE transmission.

The mechanics of containment

My second example illustrating the FORCE representation is of a simple hidden mechanism. Prajna Das Gupta and I have found that six-month-old infants apparently make a mechanical construal of invisible displacement by container (Leslie & Das Gupta, submitted). Infants apparently expect an object that has been entrained by an inverted cup to reappear in its new position when the cup is lifted up (see Figure 2). According to Piaget (1955), this is one of the last events to be understood by the infant in the course of object concept development. Although the twelve month old has solved the A$\bar{\text{B}}$ error, she has done so because these problems involve only "visible hidings": on its way to being hidden, the object's displacement is visible. Beyond twelve months, infants still have problems with hiding tasks that involve invisible displacement. This is Piaget's other "classic hiding task": A small object is enclosed by a hand; the infant then watches as the clenched hand moves under a screen. After a moment or two, the clenched hand emerges. The infant searches in the hand but finds nothing. Where is the object? Typically, infants will not go on to search under the screen until they are about 18 months old, despite the fact that they have searched under screens from 8 months and solved "visible hiding" tasks with screens from 12 months.

Figure 2 about here

It is only with the solving of invisible displacement problems that Piaget credits the infant with a formed object concept. This is because, in order to understand invisible displacement, the child must represent an absent object as having an objective existence in space—something which is not required, according to Piaget, in order to solve simpler hiding tasks, such as the A$\bar{\text{B}}$ task.

Our test of invisible displacement simplifies Piaget's task in two ways. First we dispense with the final hiding screen. Having a hiding screen at the end of an invisible displacement requires the infant to understand an embedded hiding: First, the object is hidden within the container that performs the displacement and then the container itself is hidden behind a screen. It may be that this kind of nested hiding requires too much from limited processing capacity. Secondly we used visual attention measures rather than manual search. Instead of showing the infant that the object is not in the first hiding place and asking her to search for it, we show the infant that the object is not in the second "hiding" place and ask if the infant is surprised.

Figure 2 displays one of the conditions in our study. Infants were habituated to a containment event and then tested on one of two test events. The first test event depicted an impossible outcome in which the contained and entrained object fails to reappear in Position 2. We called this the Vanish Event. The other test event we called No-Object and involves the component movements of the containment event but without there being an object present on the table. If the infant simply responded to meaningless sub-components of the containment event, then the infants should either look slightly longer at the No-Object event because this shows fewer sub-components compared with the Vanish event and is thus more novel, or, on weaker assumptions, the infants should simply continue to habituate to both Vanish and No-Object. Likewise, if the infant simply viewed the containment event as a series of meaningless "displays" appearing in positions 1 and 2, then the infants should recover looking as much or more to the No-Object event as to the Vanish event. In contrast, if the infants construe the event as involving the mechanical containment and transport of a substantial object, then they will be surprised at the "impossible" Vanish event.

We added a second condition to this experiment which we called the No-Containment Condition. The infants were habituated to the No-Containment Event as shown in Figure 3. These infants received no training during habituation on the disappear-reappear contingency involved in the Containment event. We then tested their reactions to Vanish in the absence of such training. If infants simply respond to the presence or absence of an object on stage rather than to the mechanically anomalous vanishing of an object, then again they should prefer the No-Object event. Conversely, a mechanical interpretation without contingency training will lead them to prefer the anomalous Vanish event. Our results showed that even six-month-olds prefer the Vanish Event in both conditions. Figure 4 shows the looking patterns we found.

Figure 4 about here

When the infants responded to the non-appearance of the object again in Position 2 as the container was lifted up, they were presumably surprised that the object was not there sitting on the table top. Apparently, infants do not expect the inverted container to support the object—for the object to be stuck up inside the cup. Though the solid sides of the container can and do entrain the object sideways, the inverted cup's aperture cannot entrain the contained object upward. An aperture, of course, can neither bear nor receive nor transmit FORCE. The FORCE representation provides a natural account of how infants can make useful assumptions regarding and rapidly learn about simple mechanisms, such as the inverted container. In each case, learning about elementary mechanisms will involve relating the geometry of solid 3-dimensional objects to their FORCE dynamical properties as they come into contact with other bodies.

Mechanical Agency

My third example concerns the infant's appreciation of mechanical Agency. In terms of the FORCE representation, Agents are that class of physical object that possess an *internal and renewable* source of FORCE. Because objects move as a result of FORCE, and because Agents have an internal and renewable source of FORCE, Agents are free to move about on their own—what Premack (1991) calls "self-propelledness." Mere physical objects, on the other hand, lack an internal and renewable source of FORCE and therefore move only as a result of receiving FORCE externally from Agents or from other objects that bear FORCE transiently. This simple FORCE dynamical assumption, relating patterns of motion to the force properties of the objects exhibiting the patterns, provides a powerful learning mechanism for infants.

Some years ago, I carried out a series of experiments which probed whether infants perceive a mechanical interaction between a hand and another physical object as the hand picks up the object (Leslie, 1982b, 1984a). In these experiments, infants were shown to be sensitive to the spatial relationship between the hand and object, but only when the hand and object moved together. Furthermore, an analogous spatial relationship between two physical objects, neither of which were Agents but which moved together as the hand and object had done, did not produce recovery of interest. Figure 5 illustrates the stimuli used. The infants apparently considered the contact relation to be significant only in the case of active hands. Presumably this was because only in the case of the active hand does the spatial contact mediates a mechanical relation, that is, allow the transmission of FORCE.

Consider the following. A hand moves around a lot on its own. A mere physical object does not move around a lot on its own. Suppose the infant registers these two spatiotemporal patterns of motion. It does not follow from these spatiotemporal patterns alone that, when the infant sees the hand and object moving together—with *symmetrical* spatiotemporal patterns—that it must of necessity be the hand that is picking up the object rather than the object that is pushing the hand backwards. The mere fact of previously associated spatiotemporal patterns should not, in and of itself, produce a perception of *asymmetrical* mechanical roles in an event in which the hand and object move together. However, an asymmetrical interpretation of mechanical roles does follow if one applies the FORCE representation: Because the hand moves around a lot on its own, it provides ample evidence for **ToBy** that the hand has an internal and renewable source of FORCE. The passivity of the object, by contrast, indicates to **ToBy** that the object depends on external sources of FORCE. Having arrived at a FORCE dynamical description of hands as Agents, it does follow that when hands and objects move together it will be the hand that transmits FORCE and the object that receives. From this mechanical point of view, it will be the hand which picks up and pulls the object rather than the object which pushes the hand backwards. The FORCE dynamical interpretation leads the infant naturally to attend to the contact relation, since spatial contact is required for the transmission of FORCE and, more generally, for the mediation of all FORCE dynamical relationships.

Further evidence for the FORCE representation

In order to indicate the possible scope of the FORCE representation, I shall briefly point to a range of other mechanical phenomena understood by infants. Many of the findings of Baillargeon (see e.g., this volume) and of Spelke (see also this volume) fall under the rubric of the FORCE representation.

The FORCE representation plays a role in defining the mechanical object. There is an important psychological distinction between mechanical object and the "purely visual object" (see Leslie [in press] for discussion of this distinction). The purely visual object is the output of visual processes that recognise objects visually by shape, that is, according to the disposition of their surfaces or shape parts in space (Marr, 1982). Purely visual processes probably do not make explicit any information about the mechanical properties of objects (see the discussion of the case of solidity in Leslie [1988b]). Vision is confined to representing the visible properties of surfaces.

The mechanical object, however, is constituted by properties of its interior as much as by the visible properties of its surfaces.

From the point of view of the FORCE representation, the most basic property of a physical body is its *hardness*: to be a body and to enter into the system of FORCE dynamical relationships, an entity must have a non-zero degree of hardness. A simple thought experiment suggests that if something entirely lacks hardness, our commonsense intuition is that it is not, by virtue of that fact alone, a physical body. A purely visual object, on the other hand, does not have to possess any degree of hardness. Imagine, if you will, entering a room in which a superb 3-D hologram of a cup is displayed sitting upon a table. As long as one is restricted to examining it visually, one is convinced it is a cup. Those visual processes that recognise objects by shape, immediately recognise this apparition as a cup. However, when one tries to grasp it, one's fingers pass through the visible surfaces with no resistance whatsoever. Now one will say that, after all, there is “nothing there”. However, the apparition will continue to be a perfectly good visual object.

If non-zero *hardness* is the prerequisite for something to count commonsensically as a physical body, it is because this condition must be met, if an entity is to enter into the system of FORCE relations. Without some degree of hardness, an entity can have no contact mechanical interactions with anything else.

The innate principles of objecthood that Spelke has identified (e.g., Spelke, 1988; Spelke *et al*., 1992) can be seen against the background of these remarks. Thus, the mechanical object constituted in terms of *boundedness* resists intrusion along its boundaries by other FORCEful objects to the extent that it is *hard* along its boundaries. It resists deformation by FORCEful objects to the extent it is rigid, and rebounds from deformation by FORCEful objects to the extent it is elastic or compressible. In a contrast case, a mass of sand, having a non-zero degree of hardness and thus having mechanical properties and qualifying as a body, does not qualify as a (count) object because it has no hard boundaries.

A mechanical object is *cohesive* to the extent it remains a continuous whole when subjected to FORCE. Whereas a visual shape description will make explicit, for example, that a pyramid is on top of a cube forming a continuous *shape*, it will not reveal whether the pyramid is *attached* to the cube forming a single object. Attachment is a mechanical relation: whether two entities are part of a single object or form two distinct objects depends upon the FORCE required to separate them. More elaborate notions of attachment, breakage, fixing, and so on can be developed within this representational system allowing the construction of a mechanical object catalogue (cf. Corballis, 1992, on the human capacity to create artefacts). Finally, *solidity* appears to be related to hardness in a general sense, as well as having a more particular sense of *being filled*, that is, having a hard interior. The solid object resists occupation by another FORCEful object unless its inside is empty and there is an aperture in its outer surface through which the other object can pass into its empty interior. Objects like this can function as mechanical containers.

The FORCE representation allows objects to be described as simple mechanisms. Perhaps the simplest “mechanism” of all is the movement of an object. The blocking of one object by another (Baillargeon, 1986; Baillargeon, Spelke & Wasserman, 1985; Spelke *et al*., 1992) illustrates the *resisting* of FORCE. The launching of one object by another (Leslie & Keeble, 1987) illustrates *transmission* and *reception* of FORCE. Transport by mechanical containment illustrates *resistance* and *entraining* (Leslie & Das Gupta, submitted). Even the static relation of support studied by Needham & Baillargeon (1993) may illustrate an implied FORCE in the supported object which is resisted through contact with the support.

Finally, there is a particular class of mechanical object which is distinguished by having its own internal and renewable source of FORCE, the class of Agents. This characterisation of Agents within the FORCE representation determines that Agents will be the source of FORCE in interactions with mere objects, that is, determines the direction of causation: a hand will therefore be perceived as playing the active mechanical role in events such as pick-up (Leslie, 1984a). Agents enter ubiquitously into FORCE mechanical descriptions by moving freely, and by being a source of support, blocking, pushing, transport, and so forth. This notion of the mechanical Agent falls out naturally from the FORCE representation.

It is beyond the scope of this chapter to compare the FORCE representation with the mechanical theories that children and adults come to hold as a result of prolonged reflection or formal instruction. Carey (e.g. Carey, 1986) has described some of the differences between the commonsense construals of physical phenomena children make and scientific accounts of the same phenomena, as well as some of the difficulties children have responding to instruction in these scientific ideas. It will be important no doubt to distinguish between core representational systems and the commonsense theories that later develop around this core, and that yet later resist replacement by scientific ideas. The later "peripheral" developments, while perhaps reflecting aspects of the core, will tend to focus on the areas where the core notions cast least light. For example, McCloskey (1983) discusses parallels between 'naive' adult notions and medieval impetus theories. The impetus theory of motion stated that an object that has been set in motion has acquired a FORCE that keeps it moving and then gradually dissipates. Many adults today apparently hold some version of this view. My guess is that the parallels are not wholly relevant to the FORCE representation. This is because the domain of the FORCE representation is contact mechanical interactions between objects, whereas the central problems for impetus theory are the behaviour of objects in "free flight": projectiles, rising and falling objects, and so on. One might speculate that core mechanics focuses rather narrowly on the moment of contact: the FORCE dynamics of a motion are specified less and less definitely the further away from the moment of contact the objects travel.

In summary of this section: I postulate a specialised adapted learning mechanism, **ToBy**, which, deploying the FORCE representational system, makes explicit information about the mechanical properties of objects and events. The notion of the mechanical Agent springs naturally from this representational system, as do representations of the mechanical interactions between Agents and mere physical objects. What the FORCE representation will not do is capture the next two levels of understanding Agency.

ToMM and Intentional Agency

If we stuck to the mechanical level, we would fail to notice the other distinctive properties of Agents, for Agents are not merely involved in mechanical events. Agents also *act* in pursuit of goals. Tracking the actional properties of Agents means relating Agents to the environment at a distance in time and at a distance in space. This contrast sharply with the contact mechanical concerns of **ToBy**. The FORCE representation focuses **ToBy** on contiguous relationships in time and space. But when an Agent acts in pursuit of a goal, the goal state is a state of affairs that is not yet come about; it is, as it were, "at a little distance in time." The goal in fact may never be realised in which case the goal state remains, as it were, “at a little distance in time.” Agents act not only in pursuit of goals, they also re-act to the environment through perception. Such re-action to the environment takes place at a distance in space. Agents act and re-act in relation to circumstances at a distance in space and time. This is the fundamental principle of the second level or sub-theory of Agency: Actional Agency.

This mode of attending to Agents—that is, at the level of action—creates a quite different information processing task from tracking the contiguous relations of mechanics. It seems reasonable to assume that this work is carried out by a quite different brain mechanism from **ToBy**. Accordingly, in my model, actional Agency is tracked by a mechanism which I call **ToMM**, for “Theory of Mind Mechanism.” The properties of Agents which are tracked by **ToMM** also differ in another quite fundamental way from those properties tracked by **ToBy**. **ToMM** is concerned with the *intentional* properties of Agents. Whereas objects and physical events are simply *in* the world, the intentional states of Agents are *about* the world. **ToMM**'s task is to provide a spontaneous grasp of such states: to provide a system for representing how meaning enters into the determination of behaviour. Whereas **ToBy** provides an object-centered description of a given scenario, **ToMM** provides an intentional or Agent-centered description.

The actions of Agents are not simply reducible to a mechanical pattern of motion any more than mechanical patterns of motion are simply reducible to spatiotemporal patterns. The intentional level reflects the fact that the behaviour of Agents is *about* circumstances that are at a distance from the behaviour itself.[3] In fact, my model recognises two sub-types of intentionality. Accordingly, I shall talk about **ToMM** $system_1$ and **ToMM** $system_2$. **ToMM** $system_1$ has the job of making explicit information about the state of affairs that the Agent is *acting* to bring about or *trying* to bring about. **ToMM** $system_2$ has the job of making explicit the *attitude* that the Agent takes towards the truth of a given piece of *information*. I shall be brief in my exposition of **ToMM**. For further discussion of **ToMM** $system_1$ see Leslie (in press) and for discussion of **ToMM** $system_2$ see Leslie (1987b, submitted; Leslie & Thaiss, 1992).

My model provides for two sub-theories of intentional Agency: Agent and Action, and Agent and Attitude. The Action sub-theory illustrates the second causal paradigm, that of “teleological causality”. The Attitude sub-theory introduces the third and final causal paradigm, “psychological causality”. In the course of development, these sub-theories are distinct and hierarchically organised. $System_1$ provides the input descriptions of Agent's goal-directed actions that $system_2$ requires to carry out its task of inferring Agent and Attitude descriptions, just as **ToBy** provided $system_1$ with the right kind of information to enable $system_1$ to infer Agent and Action descriptions.

ToMM $system_1$ makes explicit information concerning principally three things. First, that a given Agent acts in pursuit of a given goal; secondly, that a given Agent re-acts to a given aspect the distant environment through sensing; and thirdly, that two or more given Agents inter-act when the pursuit of goals mesh in given ways. I believe there is some evidence to suggest that the understanding of actional Agency begins during the first year of life, probably in the second half of the first year. This may be reflected in the patterns of interaction which infants apparently become sensitive to during this period. Such patterns of goal interaction may include

---

[3] In hitting a nail into a piece of wood my goal might be to build a house or it might be to build a boat. Whichever it is, the motion of hitting the nail is exactly the same.

helping and *harming* (cf. Premack, 1991, this volume), *refusal*, and *instrumental requests*. The $system_1$ action representation provides a natural way to capture such patterns. For example, helping occurs when the goals of one Agent's action enhance the achievement of the goals of another. Harming occurs when the goals of one Agent oppose or diminish the achievement of the goals of another. Refusal occurs when one Agent's goal involves another Agent and is blocked by the second Agent. Instrumental requests occur when a gesture is directed at another Agent in an effort to recruit that Agent to the pursuit of the requester's own goal. Elementary examples of these actions and interactions can be observed beginning in the latter half of the first year. For example, the infant refuses food or a toy, requests an object with a reach toward it and a look at mother. Also appearing during this time are reversible role interactions like give-and-take games (e.g., Bruner, 1976). The beginning of following another person's eye gaze also occurs during this period (e.g., Butterworth, 1988). It seems plausible that these developments depend upon the recognition of the actional properties of Agents.[4]

We come now to the last level of the tri-partite theory of Agency. This level tracks the *cognitive* properties of Agents. Tracking cognitive properties requires a sensitivity to the relationship between Agents and *information*. I have argued in a series of papers (e.g. Leslie, 1987, submitted; Leslie & Thaiss, 1992) that there are a set of key notions that allow the child to understand behaviour in terms of the Agent's relationships to information. The "informational relations" represent the Agent as taking or holding an attitude to the truth of a proposition, e.g. *believing* this or that. Because the "information" may in fact only be a fiction, the child gains the capacity to do something that at first sight appears almost useless: she can begin to understand the role of fictional circumstances in producing Agents' behaviour.

It is, of course, absurd to explain the actual behaviour of mere physical objects as issuing from circumstances which one knows are purely fictional. Yet this is a hallmark pattern of explanation in so-called "theory of mind." For example, John jumped into the shop doorway. Why did he do this? To avoid the rain. Was it raining? No! John merely thought it was raining. Here the fictional circumstance of "it is raining" is used to explain John's actual behaviour though it could never be used to explain why John was wet. Such "fictional causes" pose peculiar problems for an information processing system whose job it is to track the behaviour of Agents. Fictional circumstances create a unique learnability problem. One cannot learn about fictional causes in the "ordinary" way, by looking at the circumstances, correlating circumstances with outcomes and so on, if the circumstances do not exist.

This learnability problem is partly solved by a particular data structure that is proprietary to **ToMM** $system_2$. This is the representational system which I have dubbed the "meta-representation" (e.g., Leslie, 1987). The metarepresentation captures information necessary for expressing propositional attitude concepts. In this representational system attitudes are represented as three-place relations between an Agent, an aspect of reality, and a proposition or piece of information describing a fictional situation. The first clear evidence of the use of attitude

[4] I think the evidence is reasonably good that the great apes understand aspects of the actional properties of Agents (see e.g., Premack & Woodruff, 1978; Gomez, 1991; see also Whiten & Byrne's (xxxx) anecdotal evidence although the latter authors do not give this interpretation to the observations they report). The evidence for understanding propositional attitudes in great apes seems much flimsier in comparison.

concepts in infant development occurs in the second half of the second year of life with the emergence of the ability to pretend and to understand pretence in others (Leslie, 1987).

ToMM *system*$_2$ may begin its development in the months immediately leading up to the emergence of pretence. During this period, more sophisticated forms of communication than those seen in the first year of life become evident. For example, the instrumental pointing of the first year is supplemented by informative pointing during the first part of the second year of life. Baldwin (in press) has shown that from about 18 months of age infants take into account the focus of attention of an Agent during the ostensive definition of words. Rather than take the referent of the novel word to be the object that the infant herself is attending to at the moment the word is heard, the infant instead checks to see what object is the focus of attention for the speaker and takes *that* to be the referent of the novel word. In this the infant reveals a construal of the Agent as an object which is sensitive to, and thus can convey, information. As well as discovering *linguistic* meaning by attending to the direction of Agents' attention during ostensive acts, infants around this time become sensitive to *speaker's* meaning during pretence. Thus, when in the course of play mother hands baby a banana saying, “The telephone is ringing. It's for you,” the infant does not conclude that “telephone” is a word for bananas, but correctly interprets mother's actions as relating to mother's pretence that the banana is a telephone.[5]

Summary

The integrated cognitive systems described in this three part theory of Agency constitute a powerful learning device that lays the foundation for a formidable social intelligence as well as for a causal view of the world of considerable richness and subtlety. Three distinct forms or paradigms of causation are contained within the three representational systems deployed in the course of infancy. I suspect that many later culturally elaborated domains of knowledge and belief are based on extensions of these three paradigms and that productive tensions may arise between them when they compete to capture phenomenon.

There are also aspects of the cognition of Agency that are left out of this picture, for example, affective evaluation and moral judgement. As regards basic moral evaluative judgement (see Premack, this volume), my feeling is that this “belongs to” or at least is related to **ToMM** *system*$_1$ and the sub-theory of Actional Agency. This is suggested by, e.g., the notions of helping and harming, in particular, and by the need to evaluate goals and interactions, in general. In the syndrome of childhood autism there is a specific impairment in *system*$_2$ notions alongside a relative sparing of *system*$_1$ notions (see e.g., Leslie & Roth [1993] and Leslie [1992] for a short review). Recently, Blair (unpublished) has shown that autistic children are able to distinguish between actions that transgress moral injunctions from those that merely contravene convention. This interesting neuropsychological result fits nicely with the picture of core architecture for the cognizing of Agency that has been emerging over the last decade.

---

[5] These and related issues are discussed at length elsewhere and I refer the interested reader to Leslie (1987, 1988a, submitted), Leslie & German (in press), Leslie & Roth (1993).

QUESTION SESSION

<u>Elizabeth SPELKE</u>: I like your idea that the mechanical system for representing objects and their interactions is distinct from the visual system for representing surfaces. There are good reasons for this: one needs the visual system to represent *whatever* happens, things dissolving, things interpenetrating. But one needs the other system to underlie inferences about what will happen/what should have happened out of view. But does the notion of FORCE in itself help in supporting predictions/inferences about interactions among inanimate objects? (A) Why does the infant see the moving object as the Agent instead of the stationary object? (B) Why do infants appear to make some predictions about object motion (eg. solidity) but not others (eg. inertia)? Will the FORCE notion ultimately explain this, or does it need to be supplemented by further principles and constraints?

<u>LESLIE</u>: The FORCE notion can draw together and make sense of otherwise disparate phenomena. Objects which move are (typically) seen as bearing FORCE, those that do not are (typically) seen as lacking FORCE. For a FORCE transaction to occur there must be contact between objects. This provides powerful assumptions for learning about object interactions. For example, in launching, the moving object must be the cause because it is the bearer and therefore the transmitter of FORCE on contact with the stationary, FORCEless object. The respective roles that the different objects play in this encounter follow from the FORCE representation. (Notice that I do not say that the moving ball in launching is an Agent—it is momentarily the cause but it is a mere physical object throughout because it lacks an internal and renewable source of FORCE.) The only alternatives to this sort of account that I know of rely upon simply describing the spatiotemporal patterning of motion or appeal, à la Hume, to statistical properties of the event. None of these alternatives, as far as I can see, account for or do justice to the empirical data. For example, they do not tell us why infants see launching as an event with a particular causal direction over and above its spatiotemporal direction.

These simple FORCE assumptions also extend to mechanical Agents. (Let me re-emphasise that "mechanical Agent" means an Agent described at the mechanical level—alternatively, an entity which is an Agent by virtue of its mechanical properties—and not simply any physical object that happens momentarily to be mechanically active). Because Agents move around freely and without the external acquisition of FORCE, they must have an internal and renewable source of FORCE. Assuming the infant has sufficient exposure to a given Agent, this is the only description compatible with the FORCE representation. It also fits nicely with **ToBy**'s concern with internal properties, to which I drew attention. So it seems to me that your first question is answered decisively by the FORCE notion.

Why are inertia (and possibly gravity effects and acceleration) harder for infants? My guess is that this stems from the FORCE dynamical interest in contact. If force can only be carried by bodies, then it can only pass between bodies through contact. A property of object motion such as inertia may not be construed as a unitary property of motion but instead partly in terms of mechanical role—e.g. whether one object resists another in an interaction—while other aspects of our cultural notion of inertia may not fall under the FORCE representation at all. If so, this would be an important empirical finding with regard to our core representations. ToBy may have no in-built specifications for "free-flight" motion patterns (gravity, acceleration, etc.), and these patterns may simply have to be learnt by gradually internalising spatiotemporal analogues. In this connection one might raise the case of *support*. Some of your own results suggest that infants lack an appreciation of support relations, while other work by Renée Baillargeon and collaborators (e.g., Needham & Baillargeon, 1993) suggests that infants can appreciate support. If ToBy has little or no specification for "free-flight" patterns then an object stopping suddenly in mid-flight might not seem surprising even though it lacks visible support. Perhaps this is no more surprising than other events in which objects suddenly start to move or suddenly stop moving without there being a visible external cause. In Baillargeon's tests of support, there are contact mechanical relations between one object and another: an object is pushed over the edge of a support. It may be that these contact mechanical relations engage the FORCE representation in a way in which "free-flight" does not.

Other findings coming out of Renée Baillargeon's lab may also point to the sort of contribution that network tuning may make to the infant's reasoning about the physical world. It seems that infants apply fairly general mechanical principles to understanding elementary mechanisms but are relatively poor at problems which demand *metrical* reasoning. So whereas the baby assumes that one object cannot pass through another object, the baby is unsure where exactly on its trajectory the first object will contact the other (see Baillargeon, 1993; this volume). This sort of metrical information probably relies on analogue representations, while mechanical role (FORCE dynamical) representations are well suited to predicate-argument representation. If this is so, then there are interesting implications for the internal architecture of **ToBy**, implying a close connection between two different kinds of architecture within a single system. Something similar appears to be true within the language system, within, in fact, a tiny sub-part of language concerned with the morphophonemics of past tenses. Pinker & Prince (unpublished) have argued that regular and irregular past tenses are handled by systems with quite different architectures, the one rule based and structure dependent, the other associative and "metrical".

Frank KIEL: Is the idealised representation provided by the FORCE system truly ideal in its abstraction nature such that it would induce the infant to actually prefer the pure idealised launching *more* than the imperfect, friction-full sorts of launching so common in the real world? Indeed, it could be argued that aside from infants who hang around billiard parlours, most have never seen anything approaching an idealised launching event. If you think that infants would not prefer the idealised over the more natural one, what then do you mean by saying the FORCE representation is idealised?

Does the FORCE system enable understanding of things like a container or Spelke's primitive notions or does it take those as arguments and expand on them? One can have two-dimensional notions of containment, and the notion of spatiotemporal continuity seems to be perhaps given new implications by FORCE but not explicated or understood in such terms.

LESLIE: The effect of the ubiquitous friction on real world motions raises interesting questions for the representation of FORCE. For example, perhaps it appears to infants that mere objects suffer a spontaneous loss of FORCE in contrast to Agents who move freely with a renewable (internal) source of FORCE. Whether or not spontaneous loss of FORCE is construed by infants I have no idea. It is however interesting that Michotte found that launching events where the second object slowed down after its initial impulse seemed to produce a stronger causal impression in adults. I am not sure that I said that the FORCE representation is "idealised." I did argue that the FORCE representation could explain why a launching event should appear as a "perfect instance of cause and effect" quoting from Hume, whereas Hume was not able to offer any explanation for this at all.

With regard to your second set of questions, I do believe that the FORCE representation plays an important role in understanding containment *qua* mechanism. It is true that one can have two-dimensional containment like a circle with a cross in the middle. But here the containment is purely spatial without any mechanical significance whatsoever (see the discussion of this point in Leslie, in press). When however the containment is provided by a three-dimensional object, then there are mechanical implications which have to be understood in terms of FORCE and related notions (see e.g., my discussion earlier of *hardness*). For example, a 3-dimensional container can entrain or transport another object placed within it, while a 2-dimensional container cannot (do not confuse the 3-dimensional piece of paper with the 2-dimensional circle and cross upon its surface!). Leslie and Das Gupta (submitted) have evidence that infants as young as six months understand some of the mechanical implications of such containment, as I outlined earlier.

Spatiotemporal continuity is not likely to be explicated in terms of FORCE nor should we expect it to be. My assumption is that spatiotemporal properties are described prior to the operation of **ToBy** and indeed these descriptions form **ToBy**'s main input (see Leslie, in press). What you may have in mind is spatiotemporal continuity in relation to object identity. It seems unlikely that the FORCE representation is fundamental to object identity (in the sense of numerical identity). However, there are cases where FORCE considerations play some role; for example, how identity is determined is affected by whether or not an entity is a countable object or a mass. It is necessary to describe the FORCE properties of an entity, e.g., its cohesion, to determine whether or not it is countable or simply an incohesive mass. One way of doing this might be to apply minimal FORCE to the entity by, for example, poking it gently with one's finger. However, I do not wish my message to be that somehow "everything" reduces to the FORCE representation.

Susan CAREY: There are two distinct senses of "explicit": (A) on the surface of the representation that is the output of some stage of processing (B) available to the central reasoning process, ie. on the surface of the output of the last stage of processing before reasoning. Which do you mean? To focus this question: would you predict that eighteen-month-olds, for example, would be able to make explicit judgements on the basis of the output of ToBy (for example, verbally expressed predictions or predictions by pointing to a photograph choice of what will be found behind a screen in a Spelke/Baillargeon scenario)?

LESLIE: I was using "explicit" in the first sense, that is to say, in the sense following Marr (1982), of information being foregrounded by a representational system. This use of "explicit" leaves open as an empirical question whether or not the output of ToBy is explicit in the second sense, eg. available to a given central reasoning process at a given age, in a given task, under given circumstances, etc. It seems to me that such questions will turn out to be quite complex. For example, is central reasoning is a single unstructured process? I doubt it. I think that **ToBy-**

ToMM together function as a kind of interface system between input and more central processes (cf. Premack's pertinent remarks about “interpretation” processes [Premack, 1991]). If, instead of asking about verbal prediction at eighteen or twenty-four months, your question had asked about manual search responses at twelve-months, one’s answer would be that the action planning system either has limited access to **ToBy**’s output or is not always able to make good use of **ToBy**’s output. All these questions about knowledge versus ability in development can only receive clear answers in relation to a blue-print for cognitive architecture. I have made related arguments in connection with three-year-olds’ difficulties with solving certain false belief problems (e.g., Leslie, submitted; Leslie & Roth, 1993; Leslie & Thaiss, 1992). There are limits to the attempt to describe "what the child knows" because neither the "child" nor "knowing" are psychologically indivisible, and to that extent are not the appropriate entities for psychology study.

In short, the theory of **ToBy** leaves it as an open question whether or not and under what circumstances children can verbally access the output of **ToBy**. Other views in the study of development have stressed, indeed insisted on, the homogeneity of cognitive architecture. Such views would, as a matter of principle, conclude that **ToBy**’s knowledge is continuous with central processes and encyclopedic knowledge. However, as far as I am concerned, these are empirical questions.

<u>Leonard TALMY</u>: Your use of Michotte’s "launching effect" in your research raises the question of why this effect is thought of as "launching", which refers to the later occurring motion of the second display dot, instead of as something like "stopping dead", which would refer to the earlier occurring motion of the first display dot. One can posit as a reason for this a systematic perceptual bias - both among subjects and among the professional psychologists who name the effect - to attend to the later event, the result, preferentially more than to the earlier event, the cause.

An explanation for this asymmetry in the distribution of attention in perception is a parallel asymmetry that can be observed encoded in the structure of language as a recent paper of mine, "The windowing of attention in language", describes. In referring to a causal sequence that is initiated by a volitional Agent and that ends with the desired result, there is a universal tendency among languages to favour a construction which explicitly represents only the initiator and that final resulting sub-event. For example, if I as an initiating Agent want to break a vase, and effectuate this by bending down, lifting a rock, swinging the rock, launching the rock, propelling the rock through the air, where the rock then hurtles through the air, impacts with the vase, and then the vase breaks, this entire sequence can be represented in English as "I broke the vase", which names solely the initiating Agent, "I", and the final resulting sub-event, "the vase broke". The point is that all the intermediary causal sub-events are omitted from explicit mention. Of these intermediary sub-events, the next most favoured one is the penultimate sub-event, that is, the immediate cause of the final event, which in some languages must be explicitly mentioned, while in others it is available for optional mention. An example of the latter is English, with its ‘by’ clause. So, for example, one can say "I broke the vase by hitting it with a rock", but any earlier sub-event in the entire causal sequence cannot be readily expressed in the ‘by’ clause. For example, it is impossible to say "I broke the vase* by bending down/by picking up a rock/by swinging a rock/by propelling a rock". In being structured in this way, language would seem to be reflecting general cognitive structure with respect to how attention is distributed over Agentively effectuated results.

In general, the intuition is that once one has proficiency, one's foregrounded attention is almost entirely on the final sub-event—the desired result or effect—and minimally, or only backgroundedly, on any of the intervening causal sub-events. The reason that such a distribution may have evolved into this pattern is that a volitional Agent's intended goal or result can be treated psychologically as an invariant held across the variation of the range and variety of means that might be marshalled to accomplish the intended result.

LESLIE: An alternative account of the attentional bias in regard to the launching event might be the principle "attend to the FORCE". Given ToBy's notion of FORCE and the resulting asymmetrical FORCE transaction in launching, we have a bias to follow the progress of the FORCE as it is carried away by the second object. With regard to your very interesting observations on language structure, I am attracted to the idea that there has been a co-evolution of grammatical (e.g., verb-argument) structure and the representational systems employed by **ToBy** and **ToMM**. Pinker and Bloom (1990), for example, have made parallel arguments. If I am right in understanding your examples, then where there is a volitional Agent, the Agent can be mentioned as initiator of an extended sequence; where there is no volitional Agent, then only a direct cause (ie. the penultimate event in the sequence) can be mentioned. Thus if I (a volitional Agent) detonated a dynamite charge sending a huge boulder rolling down a mountain which smashed into a dam releasing large amounts of water which then flooded Jim's house, I can readily say, "I flooded Jim's house". However, if the huge boulder in question had been sent down the mountain by the eruption of a volcano, it would be odd to say, "The volcano flooded Jim's house". I would not be surprised if such facts about language are related to the different "conceptual" contributions of **ToBy** and **ToMM**, in particular to the contrast between the contiguous relations of contact mechanics and the "at a distance" action of goal-directed Agency (see discussion p. 17 above). For example, in the case of **ToBy**'s contact mechanical FORCE transactions the focus is necessarily on immediate cause. In contrast, the contribution of **ToMM** is to focus attention on the "at a distance" link between Agent and goal (what you called "desired result").

If language has evolved (in part) to express and communicate efficiently the conceptual structures generated by **ToBy** and **ToMM**, then we should expect there to be inter-relationships of this sort. Basic aspects of verb-argument structure appear well suited to express **ToBy**'s mechanical relationships (and less well suited, for example, to represent the appearance of individual faces). **ToMM** ($system_1$) requires that the basic "mechanical" sentence structures be extended to allow mention of the Agent in its typically "distant" goal-directed, initiating role. Verb-argument structure then requires one more extension, this time to reflect **ToMM** $system_2$, where one of the arguments is an entire sentence. With this extension, the metarepresentation can now be expressed linguistically by means of verbs of attitude and argument. Together with an object catalogue and a store of nouns, these systems form a major part of core conceptualisation that is in place by the third birthday.

<u>Dennis HILTON</u>: In the case of the hand moving the object (entraining) the perception of causality may be affected by the orienting response. We know from Michotte that where the eye is fixated affects the perception of causality. Could an explanation of seeing the hand as the cause simply be that we orient to what is moving, track it with our eye and therefore see it as causal? No attribution of "intentionality" would be needed.

LESLIE: I need to make a couple of clarifications in response to your points. I did not argue in the case of my hand experiments that an attribution of "intentionality" was implied. Instead I used that as evidence to support the apprehension of *mechanical* Agency by an infantile **ToBy**. In describing mechanical Agency, no intentionality is attributed. On my modular assumptions, that is the job of the next stage of representation which is carried out by **ToMM**. As regards your suggestion of a role for the orienting response, it is not at all clear why tracking with the eyes would make you see something as a cause. Furthermore it leads to paradox in the case of the launching event, where tracking the second object as it moves away would lead one to perceive it somehow as the cause rather than the effect.

The point I tried to make with regard to my hand experiments was that simply tracking a moving object would not carry implications for any particular mechanical role when that object is in interaction with another object such that they move together. If they move together, presumably they are tracked together. So one of the events the infant saw began with a hand grasping an object and both are stationary. The hand and the object then move off together. The equivalent pattern is shown where the hand and object are slightly out of contact. In cases like these, the infant sees the hand and the object moving with the same spatiotemporal pattern. The question why prior perception of the hand moving freely and on its own should lead to the representation now and the inference now that, when the hand and the object are moving together, that it is the hand that is the cause and not the object. My suggestion for this is that a prior perception of the hand moving freely on its own has already led to a certain "conclusion" or rather to a certain FORCE representation, namely of the hand as a mechanical Agent (that is, as an entity which is an Agent by virtue of its enduring mechanical properties). Agency in this sense is a property of the hand and *consequently* of particular motions the hand undertakes. This results in a particular directional FORCE representation when a hand and object move together, namely, the hand picks up the object.

Michael MORRIS: You propose that a concept of "FORCE" is primitive in causal cognition about the physical domain rather than, say, "transfer of kinetic energy". A virtue of this more abstract concept, "FORCE", is that it can also describe people's causal intuitions about psychological and social domains, as Talmy has argued. Do you think that a common notion of "FORCE" underlies causal reasoning in physical, psychological, and social domains? Or, do you think that notions of psychological or social "FORCE" are mere metaphors from the physical domain?

LESLIE: I do not think that the notion "FORCE" is so abstract that *exactly* this notion simply shows up again in the folk psychological domain. To that extent, what Talmy has drawn attention to is a "metaphor". Maybe there are notions in other domains which are "descendants" of **ToBy**'s FORCE (see Carey, 1988, for a discussion of conceptual descendancy in development). If so, there might be two different kinds of descendancy at work. One type of descendancy might arise in the course of the evolution of core architecture where neural circuits which have provided solutions to old problems are duplicated and adapted to solve new problems. Adaptation might introduce new characteristics. For example, one might think here of a descendancy between **ToBy**'s mechanical Agent who, with its own internal resources, moves freely producing contact effects on objects and **ToMM**'s actional Agent who strives freely toward distant goals.

The other kind of descendancy might be psychogenetic, that is, arise through psychological processes in development. For example, perhaps the child develops biological ideas that employ something like a notion of "vital force", say, by extending the FORCE notion of a mechanical Agent into the biological domain and the array of concepts like energetic/tired, healthy/sick, alive/dead and so on. Pointing to what appear to be intriguing generalisations and calling them "metaphors" or "conceptual descendants" or whatever is fine; however, we are a very long way from being able to construct explanatory theories. But these are among the deep questions that developmental psychology can ask.

References

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. (1993) The object concept revisited: New directions in the investigation of infants' physical knowledge. In C. Granrud (ed), *Visual Perception and Cognition in Infancy*. pp.265–315. (Hillsdale, NJ: Erlbaum Associates).

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

Baldwin, D.A. (in press) Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*.

Blair, R.J.R. (unpublished) Role taking, empathy and the moral/conventional distinction. MRC Cognitive Development Unit, London.

Carey, S. (1986) Cognitive science and cognitive education. *American Psychologist*, **41**, 1123–1130.

Carey, S. (1988) Conceptual differences between children and adults. *Mind & Language*, **3**, 167–181.

Corballis, M.C. (1992) On the evolution of language and generativity. *Cognition*, **44**, 197–226.

Hume, D. (1740) *An abstract of a treatise of human nature*. Reprinted in A. Flew (Ed.), *David Hume: On human nature and the understanding*, (London: Collier-Macmillan, 1962).

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. (1992) Autism and the "Theory of Mind" module. *Current Directions in Psychological Science*, **1**, 18–21.

Leslie, A.M. (in press) **ToMM**, **ToBy**, and Agency: Core architecture and domain specificity. In L. Hirschfeld and S. Gelman (Eds.), *Mapping the Mind: Domain Specificity and Cultural Knowledge*. (New York: Cambridge University Press).

Leslie, A.M. (submitted) *Pretending* and *believing*: Issues in the theory of **ToMM**. Tech. Report #10, Rutgers University Center for Cognitive Science.

Leslie, A.M., & DasGupta, P. (submitted) Infants' understanding of a hidden mechanism: Invisible displacement. MRC Cognitive Development Unit, London.

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 (eds) S. Baron–Cohen, H. Tager-Flusberg, and D. Cohen, *Understanding other minds: Perspectives from autism*. pp. 83–111. (Oxford: Oxford University Press).

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

McCloskey, M. (1983) Intuitive physics, *Scientific American*, **248**, 122–130.

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.)

Michotte, A. (1963) *The Perception of Causality*, (Andover: Methuen).

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

Pinker, S. (1989) *Learnability and cognition:The acquisition of argument structure.* (MIT Press).

Pinker, S., & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences*, **13**, 707–784.

Pinker, S., & Prince, A. (unpublished) The nature of human concepts: Evidence from an unusual source. Unpublished MS, MIT.

Premack, D. (1990) The infant's theory of self-propelled objects. *Cognition*, **36**, 1–16.

Spelke, E.S. (1988) The origins of physical knowledge. In L. Weiskrantz (Ed.), *Thought without Language*. pp. 168–184. (Oxford:Oxford Science Publications).

Spelke, E.S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992) Origins of knowledge. *Psychological Review*, **99**, 605–632.

Talmy, L. (1988) Force dynamics in language and cognition. *Cognitive Science*, **12**, 49–100.

Figure Legends

1: Six-month-old infants were habituated to either a causal or non-causal sequence then tested on the same sequence reversed. In the non-causal sequence there is a half second delay between impact and reaction. Infants recover their attention more to the reversal of the causal sequence. (After Leslie & Keeble, 1987)

2: Infants were habituated to an event in which an opaque inverted container covers and object then entrains it to new position. This cup is then lifted up revealing the object. Infants were tested either on an "impossible" event in which the entrained object has apparently vanished, or on a new event in which there is no object throughout.

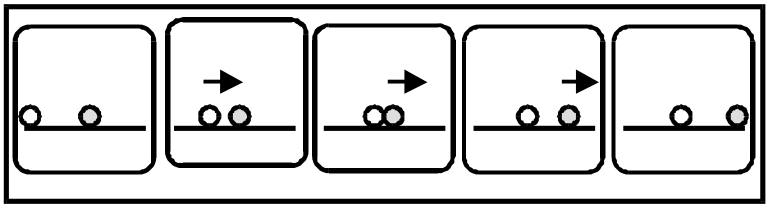
3: Infants were habituated to an event similar to that in the previous figure except that the object is not cover or entrained by the opaque container. Again the infants were surprised by the impossible test event.

4: Looking scores for infants show they were surprised at the "impossible" event Vanish event in both the containment and no containment conditions.

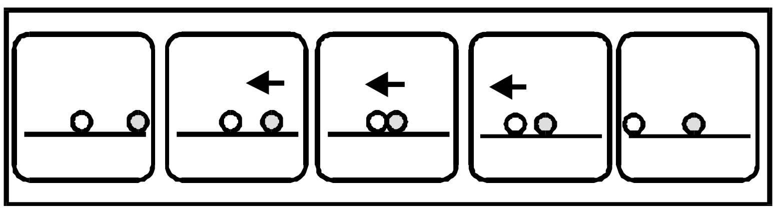
5: Reading the strip from top to bottom, infants were habituated, e.g., to a hand picking up a doll. They were then tested on a sequence showing the hand "picking up" the doll again but with a small gap between hand and object. Infants recovered attention to this change but not to a similar change when there was no hand, only another object making the same movements. (After Leslie, 1984).

Causal Sequence

Forward

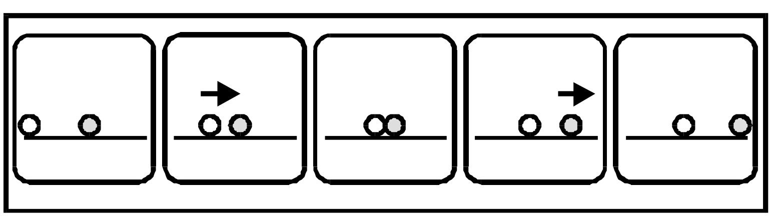


Reverse

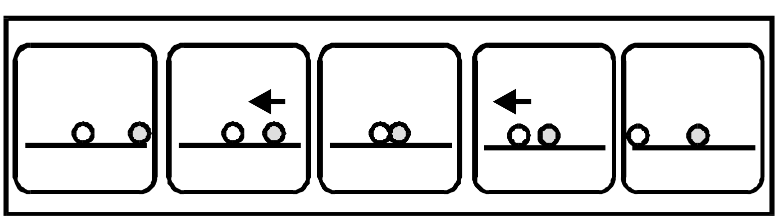


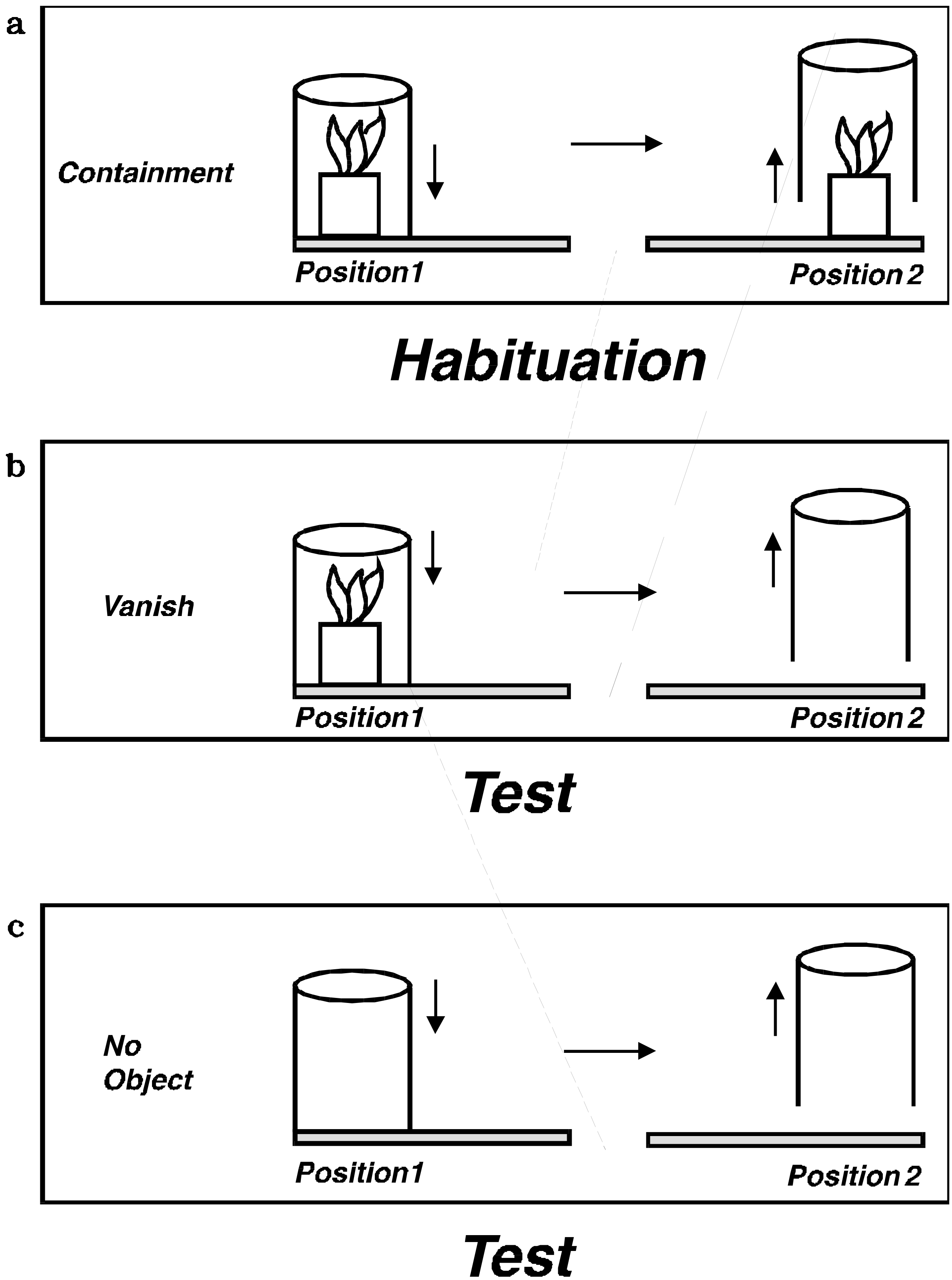
Non-Causal Sequence

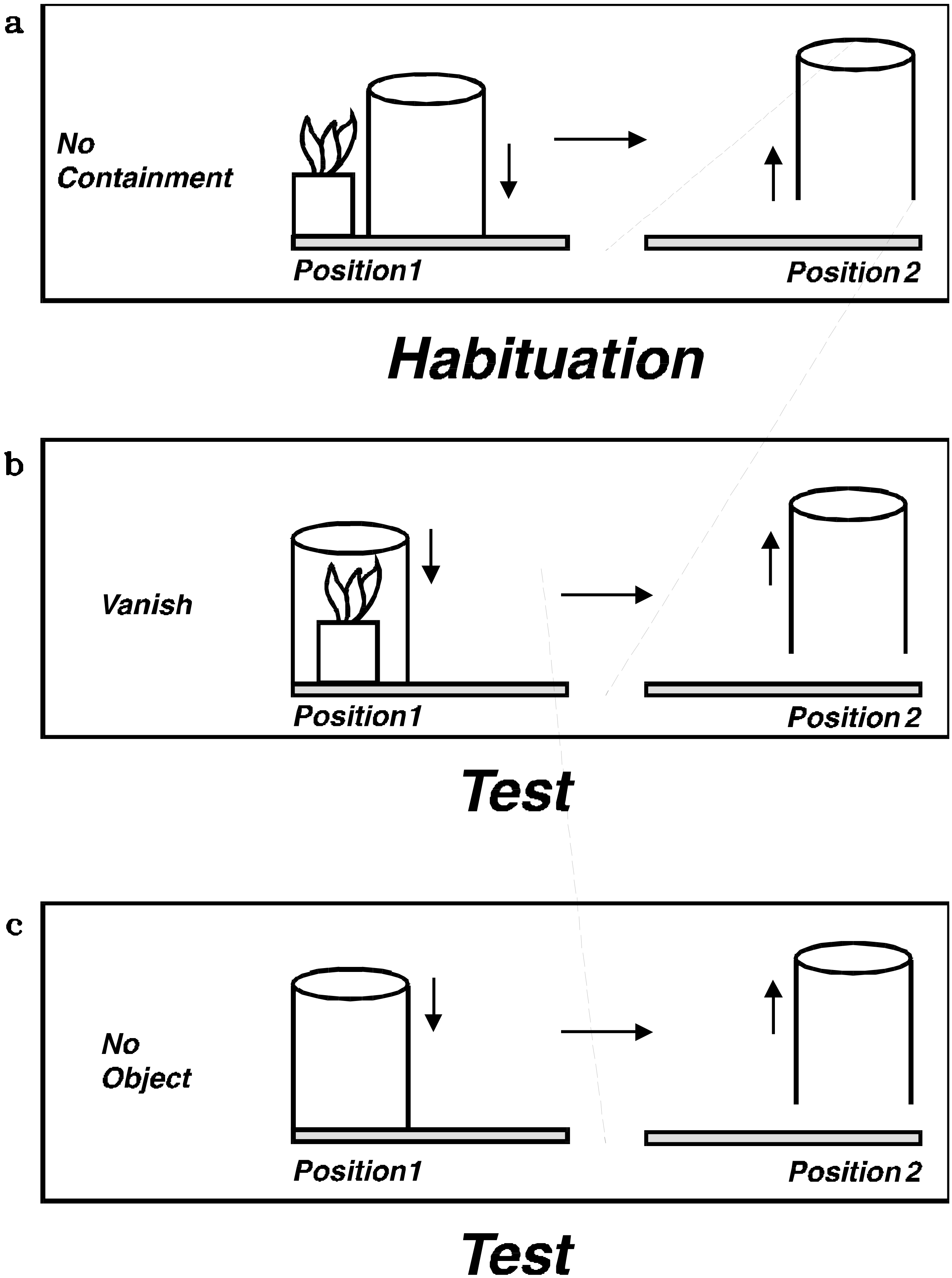
Forward

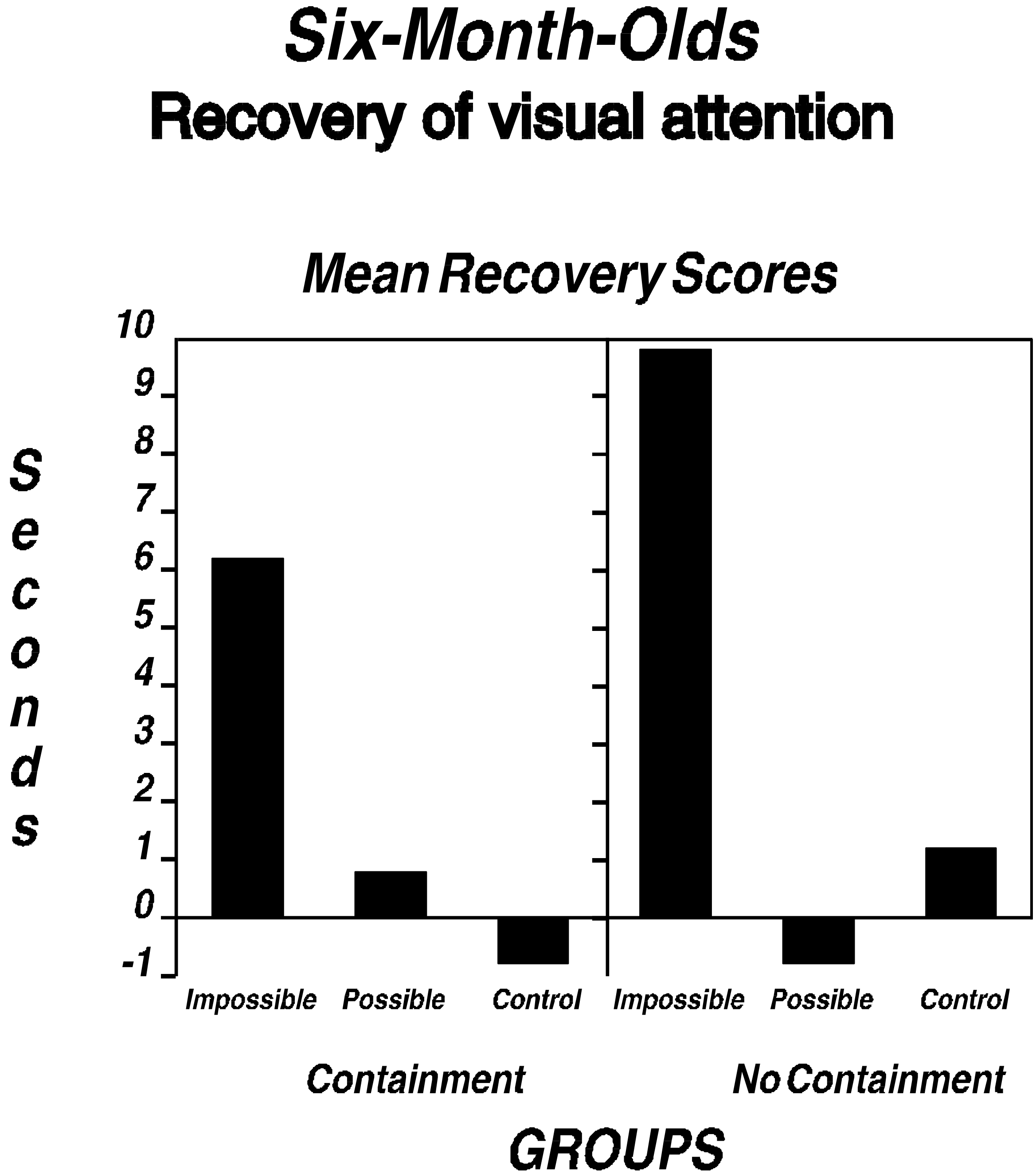


Reverse




a
Containment
Position1
Position 2
Habituation
b
Vanish
Position1
Position 2
Test
c
No
Object
Position1
Position 2
Test


a
No
Containment
Position 1
Position 2
Habituation
b
Vanish
Position 1
Position 2
Test
c
No
Object
Position 1
Position 2
Test


Six-Month-Olds
Recovery of visual attention
Mean Recovery Scores
Seconds
10
9
8
7
6
5
4
3
2
1
0
-1
Impossible
Possible
Control
Impossible
Possible
Control
Containment
No Containment
GROUPS