The infant's theory of self-propelled objects*

DAVID PREMACK
University of Pennsylvania

Received April 17, 1989, final revision accepted January 3, 1990

Abstract

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

"Theory of mind" is treated as a modular component of human social behavior and an attempt is made to find the origins of this component in the perception of the infant. According to the theory I describe here, the infant assigns a high priority to changes in motion and divides the world into two kinds of objects on the basis of this criterion: those that are and those that are not self-propelled. How the infant perceives these two kinds of objects is described by four basic assumptions. First, when the state of motion of a nonself-propelled object is changed by another object, the infant's principal hard-wired perception is causality; when a self-propelled object changes its motion without assistance from another object the infant's principal hard-wired perception is intention. Second, if two self-propelled objects are related in a special way -- a relation called the BDR sequence – the infant perceives not only intentional movement but also one object as having the goal of affecting the other object. Third, the BDR sequence has a more powerful consequence: the infant perceives that the affected object intends to reciprocate. Fourth, the infant expects that reciprocation will preserve valence (not form), where valence is formulated either as the preservation/denial of liberty, or as an aesthetic response.

Introduction

Complex social behavior is universal in humans. Like language, it occurs despite profound variation in environment, is not taught to the young, and is as characterizing of the human species as is language. On the other hand,

*Comments by Jonathan Bennett were extremely helpful, as were discussions with Pierre Jacob and Dan Sperber. It is a pleasure to express my indebtedness to Ann James Premack, with whom I have discussed the present ideas and many others over the course of a long period. The work was supported by a grant from the McDonnell Foundation. My thanks also to Ecole Polytechnique, Centre de Recherche en Epistemologie Appliquée, Paris, with which I was associated during the period in which I wrote this paper. Requests for reprints should be sent to David Premack, Psychology Department, University of Pennsylvania, 3813 Walnut Street, Philadelphia, PA 19174, U.S.A.

0010-0277/90/$5.30 © 1990, Elsevier Science Publishers B.V.

the analysis of human social behavior is at an early stage, and is beset by some of the usual problems of this stage. For example, are there modules or natural pieces into which complex social behavior can be divided, and if so what are they? A prominent contender for a modular role is that of social attribution, "theory of mind," or "folk psychology" – different names for largely the same entity. There is a core uniformity both in the acts of which this competence consists, and in the developmental unfolding of these acts in the child. In addition, there is increasing evidence for likely precursors of this competence in nonhuman primates.

On occasion, other social competences, such as "cooperation between nonkin," are singled out as a way of characterizing human social behavior, but this particular choice is unfortunate. Cooperation among nonkin is a misleading way of characterizing human social behavior; it implies that while there is no cooperation among nonkin in animals, there is cooperation among kin. Cooperation has two facets: (1) working toward a common goal and (2) (the goal having been achieved) agreement to share. Cooperation in the sense of *agreement to share* – critical to the human case – does not occur at all in animals, neither among kin nor nonkin.[1]

In this paper I have singled out "theory of mind" (Premack & Woodruff, 1978) as one of the basic components of human social behavior, and sought to find its origins in the perception of the infant. Not all of theory of mind – the goal is more modest – only one part of it, specifically *intention*. This is the main hypothesis: the perception of intention, like that of causality, is a hard-wired perception based not on repeated experience but on appropriate stimulation. In the case of causality, thanks largely to Michotte (1963), and more recently Leslie and Keeble (1987), we can specify the stimulation: basically it is temporal and spatial contiguity between appropriate events. What is the analogous stimulation in the case of intention?

---

[1]Chimpanzees hunt in groups more successfully than when hunting alone, and physically divide the meat they capture (Goodall, 1988; Boesch & Boesch, 1989); nevertheless, there is no evidence of *agreement to share*. (1) Not only is giving uncommon – the prey is divided mainly through "tolerated scrounging" – but, more important, there is no suggestion that giving is reciprocal. In reciprocal giving, donors give with the expectation that their recipient will reciprocate. A failure to give by a former recipient is censured more acutely than is a comparable failure by a neutral party (an individual who was not a former recipient). (2) There is no evidence that "payment" is proportional to contribution; that is, no evidence that the animal that captures the prey then surrenders meat more willingly to an individual that it knows participated in the hunt than to one that it knows did not. (3) There is no evidence for censure, for example, that animals which fail to give, take more than their share, take when they shouldn't take, etc. are identified as "cheaters" and driven off, excluded from subsequent hunts, etc. Indeed, one cannot speak of chimpanzees as "cheaters", for cheating presupposes violation of a prior agreement, and there is no evidence for any such agreement. What is evidently lacking in the animal are such concepts as reciprocity, equity, justice or the like that underlie an agreement to share. Cheating is thus not the concept on which to focus; it is a secondary phenomenon. Rather, one should focus on the concepts – reciprocity, equity, justice, etc. – that cheating presupposes.

Picture the infant as dividing the world into two kinds of objects: those that are and those that are not self-propelled. Self-propelled objects are those objects that can both move and stop moving without assistance from another object; nonself-propelled objects are those objects that cannot. Motion per se is not the critical parameter. The nonself-propelled object can have as its initial state either rest or motion, but in either case it will retain this state unless acted upon by another object. Change is what is critical – from rest to motion (or vice versa), one speed to another, one direction to another – and the ability of an object to execute these transitions without assistance from another object.[2] The psychophysics of the discrimination is a topic that I deal with elsewhere (Premack, in preparation).[3]

Induced changes in movement in nonself-propelled objects are what the infant perceives as causal (not all changes of course but those that carry out Michotte's temporal and spatial contiguity). On the other hand, changes in the movement of self-propelled objects are what, I suggest, the infant perceives as intentional. This then is my first argument: just as causality is the infant's principal hard-wired perception for nonself-propelled objects, so intention is its principal hard-wired perception for self-propelled objects.

Here is a more complete statement of the argument: first, the infant perceives certain properties, for example, one object is moved by the other under conditions of temporal and spatial contiguity, versus the object is self-propelled, that is, changes its state without assistance from another object; second, the infant not only perceives but also interprets, that is, the infant's perception is the input to a slightly higher-order device that has interpretation as its output. The interpretations in the two cases in question are causality and intention, respectively.

---

[2]Is it only changes in motion that produce this interpretation, or do changes in other states – color, shape, size, temperature, etc. – have a similar effect? Because motion is a conspicuous property of social objects, I have given it a privileged status. But the possible role of changes in other states is an open question.

[3]At least two aspects of the psychophysics are of special interest. First, the infant's interpretation of some cases as causal and others as intentional does not exhaust the alternatives; the infant will perceive cases that it does not interpret in either manner. Consider that the infant's perception of causality is based upon induced changes in motion, meeting criteria of temporal and spatial contiguity. And that its perception of intention is based upon noninduced changes in motion that occur at an above-threshhold rate. Changes in motion that neither meet the criteria of temporal and spatial contiguity, nor occur at a suprathreshhold rate, will not be interpreted in either fashion. Although such changes will be perceived, of course, they will not be interpreted. Second, the notion of "change" critical for the interpretation of intention is more complex than I have communicated. Suppose the stimulus given the infant is that of a "bouncing" object: the object rises, stops, falls, stops, rises, etc., thus apparently meeting criteria of change. However, if the successive cycles of the object (cycle = rise + fall) are identical, the infant may not intepret the change as intentional. Most likely, change must occur on both levels, within a cycle and between cycles.

Digression: Interpretation

The question of interpretation raises some interesting issues that I treat more fully elsewhere (e.g., Premack, in preparation), and will only mention here. Here are two such issues.

(1) Suppose that only a small proportion of perceived distinctions are inputs for interpretative devices. Most distinctions may be simply perceived, not perceived and interpreted. The distinction between perceived versus perceived + interpreted could be seen as the formal version of the casual distinction we make when speaking of what does and does not "interest" an individual. A study by Roberta Golinkoff (1975) illustrates this point. She habituated infants to a scene in which Mary gave Jane an apple. Then for half the infants she reversed the donorship: Jane now gave Mary the apple; whereas for the other half, she changed not donorship but the left/right location of Mary and Jane. The change in donorship produced significantly greater dishabituation than the change in location. I take this to suggest that whereas location is only perceived, donorship is both perceived and interpreted.

(2) Consider the implications of interpretation for species differences. Species that do not differ at all in the distinctions they can perceive may nonetheless differ substantially in those they interpret. Consider the set of all interpretations in this world. Humans probably enjoy (or suffer from) a larger proportion of this total than any other species; we engage in far more interpretation than other species. But how are the interpretations distributed over other species? Two main alternatives can be depicted by Venn diagrams, using concentric and nonconcentric circles. With concentric circles, every species' interpretations are a subset of another "more interpretive" species, and only the human or "highest" species has unique interpretations. For example, the ape's interpretations are a subset of the humans, the monkey's a subset of the ape's, etc. Alternatively, in the case of nonconcentric circles, the interpretations of different species may overlap only in part or not at all; species can have unique interpretations. Indeed, they could even assign different interpretations to the same perceptual distinctions, giving rise to a mess we would have a great deal of difficulty unraveling.

The digression has taken a depressing turn; let's close it and return to our main business.

Preference and learning as properties of self-propelled objects

If infants attribute intention to the self-propelled object, they may attribute other properties as well. For instance, preference, as well as a capacity for learning may be attributed by the infant; both are fundamental properties of intentional objects and may therefore be reflected in the intuitions of the infant.

Whether or not we find that infants attribute preference to intentional objects is likely to depend on what we choose as our example. One strong candidate is the preference a self-propelled object may have for its own kind. Will the infant too attribute this preference to the self-propelled object? The following test could answer the question.

On one side of the screen, show the infant a self-propelled object and on the other side a nonself-propelled one (randomly changing the location of the two kinds of objects from trial to trial). Arrange for a self-propelled object to appear in the center of the screen, and after a moment "join" either the self-propelled or nonself-propelled object. If infants expect self-propelled objects to prefer their own kind, they should be surprised when the self-propelled object makes an alternate choice, and should show less habituation over trials of this type than over trials in which the expected choice takes place.

One could strengthen this outcome by showing that the infant regards the preference the intentional object has for its own kind as special, at the least stronger than other kinds of preference. After placing a green object on one side of the screen, a red object on the other (both either self-, or nonself-propelled), a red/green self-propelled object will enter and "choose" between the red and green objects. If infants see the intentional object's preference for own kind as unique, then trials in which one red object chooses another object of like kind should not lead to markedly greater habituation than trials in which red objects choose green ones. The property of being intentional is, after all, far deeper than that of being red, round, large, and so forth, and this may be reflected in the intuitions of the infant.

Learning is a capacity the infant may also attribute to self-propelled objects. In this case it is not our choice of example that is important but how we exemplify learning. The main outward signs of learning are: a positive change in performance and a change that endures. This kind of change or improvement can be found either in intrinsically motivated cases – for example, an object dances better than it did, bounces better, and so forth – or in extrinsically motivated cases – for example, an object eliminates cul-de-sacs in a maze, thus achieving a goal more quickly or efficiently.

What makes the notation of learning interesting is, thus, not learning itself but what it presupposes, namely, a scale of value. The scale is evidenced in

the intrinsic case by the propensity to treat one form of movement as better than another, in the extrinsic case by the ability to recognize a "goal" and to discriminate the improved achievement of the goal from the unimproved. As to what infants might reckon as goals, consider: "freedom," the ability to move without restriction, "companionship," the proximity of another self-propelled object, "arousal," the ability of an object to increase its response rate from suboptimal to optimal.

Do infants attribute a capacity for learning to objects that they regard as intentional? The following test will answer this question. Habituate the infant to an example of learning, that is, to an object that starts at one level of performance and ascends to a higher level. Then present a spontaneous ("uncaused") loss of this achievement: The object reverts to its original level of performance. This should surprise the infant. But that alone is insufficient, for any change should produce some dishabituation. One must show that the dishabituation produced by the spontaneous dissolution of learning is unique, greater than that produced by a comparable change in a performance that was not owed to learning.

For instance, an object that began by "dancing" poorly, greatly improved its performance, and then suddenly danced poorly again, should surprise the infant. Whereas an object that began by dancing one way, changed to a quite different way but one of equal quality with the first way, and then suddenly reverted to the original way, should not surprise the infant, or should surprise it less. A close analogy is an object that "struggled" to climb up a hill and then suddenly reappeared at the bottom, in contrast to an object that went from one hill top to another before reappearing on the original hill top. The latter seems natural and needs no special explanation. The former does not: One wants to know why the object has returned to a position it sought to escape.

One can test the attribution of learning not only with the intrinsic case – for example, improved "dancing" – but also with the extrinsic, using a "goal" placed at the end of a maze. For instance, learning is demonstrated by an object that eliminates cul-de-sacs (thus attaining the goal earlier); a comparable change but one not owed to learning is demonstrated by an object that, rather than eliminating cul-de-sacs, merely "substitutes" one pattern of them for another.

In sum, infants may attribute preference and/or learning to objects they regard as intentional. While these capacities may not appear until much later in the child's development, it seems well to test for them in the infant. If the evidence is negative, we must then try to understand how these attributions do develop, for they are certainly among those that humans make about intentional objects. However, if the evidence is positive, there will be no need to apologize for having carried out an improbable test.

BDR and the perception of social goals

Now, suppose we show the infant not one but two self-propelled objects and, in addition, arrange a special relation between the two objects. Specifically, we arrange a BDR sequence, where B stands for base, D for deflection from base, and R for recovery of base. The movements constituting B, D and R can take innumerable forms and we have only begun to look at their possible composition rules (more about this in a later section). At this point, consider an example instead. Two balls, one larger than the other, appear on the screen and bounce together for, say, 5 s.; that constitutes base. Next one of the balls gets "stuck" in a virtual hole, that is, stops moving and remains immobile; that is deflection from base. Finally, the other ball "frees" the stuck one, that is, contacts the immobile ball after which it resumes moving; that is return to base.

My second argument is this: although the self-propelled object alone leads to the perception of intentional movement, adding the BDR relation leads to a further perception. Specifically, the infant perceives one object as having a goal – that of affecting the other object.[4]

To demonstrate this second argument, Dasser, Ulbaek, and Premack (1989) used habituation/dishabituation in conjunction with a role reversal paradigm. Since our subjects in this experiment were not infants but young children, 3½–5½ years old, I describe the experiment not to make claims about infants but to illustrate the test paradigm. To demonstrate that BDR leads to the perception of social goal – that one object intends to affect the other object – we compared the effect of the would-be critical BDR with the reverse, RDB. We divided the children into two groups, and after habituating one of them to BDR and the other to RDB carried out a role reversal. The ball that had been the instigator of the action, that is, "rescued" the "stuck" ball, became the recipient, and vice versa. We then showed BDR with role reversal to one group, RDB with role reversal to the other. Role reversal

[4]What bearing does the infant's ability to perceive and interpret certain relations between objects have on the special case in which the infant is itself one of the objects? The alter–alter and ego–alter cases differ in the stimulation they provide. The stimulation in the alter–alter case is purely visual, whereas in the ego–alter case it is both visual and proprioceptive, visual for the alter as before, but a combination of visual and proprioceptive for ego (i.e., the infant can both feel its own movement as well as see some part of it). Is this difference important? If the infant lacked intermodal equivalence, it would be. On the other hand, if we granted the infant intermodal equivalence, then the difference in stimulation would be unimportant, and whatever we established for the alter–alter case would hold equally for the ego–alter case. The evidence favors the positive alternative, I think. For instance, Dolgin, Spelke, and Premack (unpublished data) found that not only infant apes but even infant monkeys – 50 million years removed from *Homo sapiens*, not 5–15 as in the case of the ape – showed visual–tactual equivalence. This suggests that whatever holds for the alter–alter case will also hold for the ego–alter case.

should have a strong effect in the case of BDR, a weak effect in the case of RDB. If one perceives the interaction between the two objects as intentional, then role reversal is important and should lead to considerable recovery from habituation. But if one does not perceive the interaction as intentional, then role reversal is unimportant and should have little effect. As we predicted, role reversal produced significant dishabituation in the case of BDR, but not in the case of RDB.

Actually, the results were somewhat more complex. With young children, $3\frac{1}{2}$ years old and less, the results were exactly as described. However, with older children role reversal produced significant dishabituation with both BDR and RDB. Fortunately, we had obtained similar results some years earlier in a study concerned with the effect of violating the sequence in picture stories. Whereas presenting the pictures out of order to older children had no effect, it had a profound effect on younger ones. When asked to describe the individual pictures, younger children now said things like "he's here and he's over here" rather than "he's over here because he's afraid," omitting specifically reference to the intentional component (Poulsen, Kintsch, Kintsch, & Premack, 1979). We argued that older children already knew the story schemata; therefore they needed only the elements, being capable of putting them into order themselves. But younger children did not know the schemata, hence they needed not only the elements but also the properly ordered elements. We would now make for the BDR sequence the same argument that we made earlier for the picture story.

BDR and reciprocation

The perception of first-order social intention – that one object has as its goal affecting the other object – is not the only consequence of the BDR sequence. Indeed, it is the weaker of the two major consequences. In addition to perceiving that one object intends to affect the other, the infant also perceives that the affected object intends to reciprocate. Specifically, BDR leads the infant to expect that the initial recipient of the action will "reply", acting upon the instigator in a manner that preserves the valence (and possibly magnitude) of the act that was directed at it. That is my third argument.

Reciprocation: Preservation of valence

One could simplify the infant's theory while still granting the infant the perception of intention to reciprocate. One could say the infant expects the

initial recipient to reciprocate, applying to the instigator exactly the same act that was applied to it. In other words, a kiss for a kiss, a blow for a blow. But that theory will prove too simple, I think. It will not only underestimate the infant's computational capacity, but also miss the heart of reciprocation which devolves, I think, not around form but around valence.

The infant must be credited, I think, with the ability to perceive and code valence, thus the ability to code some movements as positive and others as negative. (Whether it can also code intensity is a secondary issue that we can leave open here.) The theory I grant the infant is this: the infant can code valence and expects reciprocation to preserve valence, not form. But what is meant by valence?

Positive valence as the maintenance of liberty

One can formulate valence along either relative or absolute lines. The two conceptions are not mutually exclusive; the infant could have both, neither or one; what it actually has must be determined, of course, by test. Here I will simply outline both conceptions and describe experiments that could help decide between them. Consider first the relative formulation; it is of interest because of its alliance with liberty.

On this treatment of valence, the infant will code as positive movements of the one object that maintain, restore, or increase the "liberty" of the other object; even as it will code as negative movements of the one object that impair the liberty of the other object. It is more than coincidence, I suggest, that Rawls (1970), in discussing his theory of justice, makes liberty his first criterion.

Let us flesh out what is meant by "liberty" using as characters the "balls" that were used in the experiment described above. Incidentally, the reader (I'm confident) has already recognized a main characteristic of the infant's theory. It is not a theory of domain-specific features, for example, a theory that cats have intentions but boxes do not. Infants may have domain-specific theories, for example, that only fractals are self-propelled, whereas nonfractals are always nonself-propelled. But if so, their domain-specific theories are not a part of the present discussion. The only assumption I make here, even remotely related to domain specificity, is that the infant can perceive objects. Since, in effect, I treat "object" as a primitive, I shall have nothing further to say about it (see, for example, Spelke, 1982 for discussion of infant's perception of object).

Here are some examples of movements that impair liberty in the one case, and aid it in another. Suppose one object repeatedly deflected the other

object from its course of action: stopped it when it was moving, induced it to move after it had stopped, changed the direction and/or speed of its motion. The infant will code this action as negative. Conversely, picture an object that was repeatedly deflected from its course of action, and another object that just as repeatedly restored the deflected object to its previous course of action. The infant will code the action as positive. In brief, the infant will code as positive those acts that restore an object to its course of action, and as negative those acts that deflect it from its course of action. When human adults were shown these cases, they tended to spontaneously label them, using such phrases as "helping" or "trying to help" in the one case, "hurting" or "trying to hurt" in the other (Dasser et al., 1989).

Notice that in order to decide whether a second object is interfering with a first object one must have some idea of the intended course of action of the first object. What would this object do if left alone? When a second object causes a first one to stop, this could be interference. On the other hand, if this is what the first object was about to do, it could be a form of assistance.

Since one can compute changes in liberty only if one can predict an object's course of action, the relative formulation makes two demands: one must be able to recognize the pattern of action that is instantiated in the base condition, and in addition be able to recognize changes in the pattern. Thus, the infant must be capable of recognizing patterns of action in the base conditions that are shown it, and, of course, those that are shown it must be appropriately simple.

To determine whether infants perceive intention to reciprocate (relative formulation) one can use the standard habituation/dishabituation paradigm used in the Dasser et al. study. The infants are shown two BDR sequences: one which they would code as positive, for example, a sequence in which, as an observing adult would say, one object "helps" another; another sequence which they code as negative, thus a sequence for which an observing adult would say one object "hurts" or interferes with another. Then they are shown these same sequences with role reversal, the previous instigators of the action now serving as recipients and vice versa. Thus, reciprocation is shown in all the sequences. But for half the infants, the reciprocation preserves the valence and for the other half it does not. The predictions made by the theory are clear. Reciprocation that violates valence should disagree with the infant's expectations, therefore producing substantial dishabituation; whereas reciprocation that preserves valence should agree with the infant's expectations, producing little dishabituation.

A simple extension of the experiment would test the hypothesis that the infant's theory of reciprocation depends on preservation of valence, not form. First, we should find that reciprocation which violates form (but preserves

valence) should produce only negligibly more dishabituation than reciprocation which preserves both form and valence. Second, a more impressive demonstration of the hypothesis is possible. Consider reciprocation that preserves form but violates valence. For example, an object that "presses" against another object when it is "stuck" in a virtual hole, thus freeing the stuck object, would be said to act positively; but an object that applied the same motion to an object on the edge of a virtual cliff, sending the object hurtling down the cliff, would be said to act negatively. A proper comparison of these cases should tell us whether the infant's theory of reciprocation is one that is based on the preservation of form or of valence.

Aesthetic theory

When valence is formulated not as a relative concept (above) but as an absolute concept, the infant is no longer required to be able to recognize changes in a predicted course of action. On the other hand, the infant must be granted a theory of aesthetics. The theory may be kept extremely simple; still it must be of sufficient power to enable the infant at least to rank order some movements on a scale of value. An example will clarify this formulation.

In the base condition, the infant is shown two "balls"; they bounce about the screen together, but not in the same way. One bounces "better" than the other – "better" meaning, of course, whatever the infant's theory of aesthetics treats as better. For example, one bounces higher (lower), faster (slower), or more (less) regularly than the other. For the appropriately endowed infant, the base condition will lead to this reaction: one object is better than the other.

In deflection from base, the second phase, the superior object acts to "assist" the other, doing so in either a weak or a strong fashion. In the weak case, it acts so as to "set an example" – mainly by persistently putting itself into the vicinity of the other object, and there "demonstrating" its skill. In the strong case, the superior object not only sets an example, it goes a step further: it "corrects" the other one, thus acting pedagogically (Premack, 1984, 1986). Although example-setting and pedagogy can, of course, also have negative versions (where the "model" acts not to assist but to impair the other one), there is no particular need to describe them here – except for the contribution it makes to this point. If one collects all the interactions that have proved useful here in clarifying the two treatments of valence (and adds a few remaining cases), one begins to find that primitive social interactions are both relatively few in number and have a more or less natural order of complexity.

In return to base, the final phase of the interaction, one can arrange for either success or failure. The less accomplished object either does or does not benefit from the assistance given it by the superior object. It now bounces more or less like the superior object, or the assistance failed and it is still back in base condition. The difference between assistance that did and did not succeed can be made the basis of an interesting question. On what criterion does the infant compute reciprocation? If it credits "effort", it will expect reciprocation even when assistance failed. On the other hand, if in the infant's eyes only achievement counts, then it will not expect reciprocation if the assistance failed. Given the general tenor of the infant's concept of intention (see next section), one may expect more behaviorism than mentalism from the infant.

Common sense and the infant's theory

The infant's concept of intention differs visibly from that of the adult or common sense. The infant's concept is an automatic reading of a perceptual input, loosely interpretable as "internally caused". For the infant, objects are intentional (or have intention) when their movements are self-propelled. Whereas for common sense, intention is not an automatic reading of a perceptual input, but an inferred state of mind based on evidence for *desire, belief*, and *planning*.

The infant's concept and that of common sense differ not only in content, however; they differ in the very subject matter that the two concern. Common sense is a theory based on real-world objects – tables, chairs, dogs, people and the like – all of which have an identity; whereas the infant's concepts are based on objects that have no identity whatsoever. What counts for the infant, as I have emphasized, is not the identity of the object but the kind of motion in which the object engages.

Despite these differences, some of the features of the infant's concept appear to have made their way into the common-sense concept, so that the "intention" of common sense is a mixed concept, having elements of both. Thus, one speaks of pulling a trigger intentionally (to contrast this with the case where, say, a push caused one to pull the trigger unintentionally), but also of killing someone intentionally. Intention is ascribed here to items belonging to distinctly different levels – a small movement of a part of the body in one case, the death of an individual in the other.

The "internally caused" movement of the finger would seem to belong to the subject matter of the infant's theory, as the premeditated death belongs to that of common sense. Premeditation or planning – the essence of com-

mon-sense intention – has, of course, no bearing on the infant's concept (infants being incapable of deciding whether the evidence does or does not bear witness of planning).

If traces of the infant's theory are still to be found in the adult – leading to the confusing application of intention to two quite different levels of action – remnants of the infant's theory are even more evident in the young child. The young child's use of intention is less discriminating than that of the adult. Certain bumps between people, as well as falls or tumbles of individual people that the adult calls accidental, the 4-year-old child calls intentional. The young child appears to be still under the infant's influence, using a criterion close to that of self-propelledness. For the adult, this criterion is inadequate; he requires additional evidence, mainly that of planning or premeditation. The infant, one might say, is as close to a behaviorist as the human gets; the child by the age of 2 years (or probably earlier – data are lacking) is already a mentalist (Astington & Gopnik, in press; Baron-Cohen, in press; Leslie, 1987; Perner, in press; Wellman, in press).

Perception, interpretation, and conception

The elegant work of Tom Shultz (1982), showing that the child can find causal relations in events that are not related by simple temporal and spatial contiguity, does not, I suggest, infirm the basic role of temporal and spatial contiguity in the infant's hard-wired perception of causality. Rather, Shultz's work underscores the difference between three levels of processing – perception, interpretation, and conception – and calls to mind the necessity of both characterizing these levels and of clarifying the relations among them.

How does the notion of causality, which in its early stages is evidently highly dependent on simple physical parameters, escape these parameters and become a more general notion? A proper answer to this question will clear up not only ontogenetic mysteries but phylogenetic ones as well. For the transition from perception to conception (for lack of better terms) is, almost certainly, uniquely human; we do not expect the ape to make this cognitive journey.

The ape, both in perception and initial (or hard-wired) interpretation, may differ little from the human: the basic habituation/dishabituation data for the two species may be difficult to tell apart. Nevertheless, the cases 5-year-old children can properly classify as causal are not ones that the ape is likely to recognize. Why? What are the devices, present in the child, lacking in the ape, that enable the one species to make the transition from perception–interpretation, with its strong dependence on physical parameters, to concep-

tion, with its relative freedom from such parameters? Although we may easily implicate language, this is merely to name a faculty rather than to explain how the faculty brings about the transition.

The adult ape, we suggested above, apparently does not show reciprocal giving; this does not mean, however, that the habituation data for the infant ape will predict the adult deficiency. The infant may show evidence for the perception of reciprocation even though the instrumental behavior of the adult does not. Disparities between these measures of behavior, or levels of processing, are not uncommon. For instance, infant apes when measured by habituation/dishabituation show recognition of sameness/difference not only of objects but also of relations – thus not only of, say, A to A but also of AA to BB (and CD to EF). However, they cannot match like relations – indeed, they cannot do this even as adults unless given special training (Premack, 1983, 1988a) – though they can readily match like objects. Finding disparities between the two levels is relatively easy; understanding the disparity is another matter.

Limitations of the infant's theory of self-propelled objects

The infant's theory contains perhaps the most essential feature of a theory of mind: in perceiving one object as having the intention of affecting another, the infant attributes to the object a representation of its intentions. Still, it is essential to recognize that the infant's theory of self-propelled objects is a highly restricted one. For example, though I claim that the infant, when shown the BDR sequence, perceives one object as having the goal of affecting the other object, I do not claim that it perceives either object as perceiving that the other object has a goal. That is, I do not claim that the infant can perceive second-order intentions. One has only to look closely at the test requirements that such a claim would impose to see that the infant is highly unlikely to be able to meet such requirements.

Similarly, I do not claim that the infant perceives nonmotivational or informational states of mind, in particular *belief* – an uncommonly strong state of mind – or even *expectancy* – a notably weaker state of mind (see Premack, 1988b for a preliminary account of the distinction between *belief* and *expectancy*). For instance, the theory credits the infant with expecting (that the recipient will reciprocate), but not with perceiving that the instigator expects the recipient to reciprocate.

Thus the theory does not grant the infant the perception of either second-order motivational states or even first-order informational states. The infant may perceive such relatively weak informational states as expectancy, but

again the reason for doubting it are the same as those noted earlier. When one lays out the habituation/dishabituation tests that would be needed to prove such a claim, and looks at the requirements such tests would impose upon the infant, one doubts the infant's ability to meet the requirements. To be sure, wisdom councils running the tests, but it is not lack of wisdom that councils doubt.

In effect, the infant's theory of self-propelled objects does not account for most of the basic components which, along with intention, make up theory of mind. This limitation on the perceptual origins of theory of mind is in no way special; it is part of a more general restriction. In proposing that both causality and intention can be traced to perceptual origins, I by no means suggest that all fundamental ideas have similar origins. This would be an extremely risky proposal. Not only is there the time-honored difficulty of identifying "fundamental ideas", but even among the ideas we are intuitively willing to grant such status, there are many for which it is not possible to construct plausible perceptual origins. I return to the example of "belief" – if there are plausible physical parameters that give rise to this interpretation, they are not self-evident.

For some purposes, it may be of interest to reformulate the contrasts we have drawn here along somewhat different lines. For instance, we contrasted induced movement – leading to the interpretation of *causality* – with self-propelled movement – leading to the interpretation of *intention*. This can be reformulated by treating intentional movement as itself a form of causality, that is, as movement that is caused *internally* rather than *externally*. Then all movement is caused, some of it externally, some of it internally. Similarly, the contrast between intentional movement in the one- and two-object cases can be redrawn. One can grant the perception of *goal* in both cases, reserving specifically social goal for the two-object case.

In contemplating these reformulations, however, one must ask whose theory they reflect – that of the infant, or of an adult reflecting on the infant's theory? In effect, which formulation is simpler? For instance, compare the original contrast between causality and intention, with the present proposal to treat all changes in motion as caused, and to distinguish between external and internal forms of causality. Although this unifying concept may be attractive to the adult, the original contrast has the advantage of simplicity. The higher-order concept may exceed the infant's capacity.

References

Astington, J.W., & Gopnik, A. (in press). Developing understanding of desire and intention. In A. Whiten (Ed.), *Natural theories of mind: The evolution, development and simulation of everyday mindreading.* Oxford: Basil Blackwell.

Baron-Cohen, B. (in press). Precursors to a theory of mind: Understanding attention in others. In A. Whiten (Ed.), *Natural theories of mind: The evolution, development and simulation of everyday mindreading.* Oxford: Basil Blackwell.

Boesch, C., & Boesch, H. (1989). Hunting behavior of wild chimpanzees in the Tai National Park. *American Journal of Physical Anthropology, 78*, 547–573.

Dasser, V., Ulbaek, I., & Premack, D. (1989). Perception of intention. *Science, 243*, 365–367.

Golinkoff, R.M. (1975). Semantic development in infants: The concept of agent and recipient. *Merrill-Palmer Quarterly, 21*, 181–193.

Goodall, J. (1988). *Chimpanzees of the Gombi Stream.* Cambridge, MA: Harvard University Press.

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

Leslie, A., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition, 25*, 265–287.

Michotte, A. (1963). *The perception of causality.* London: Methuen.

Perner, J. (in press). *Towards understanding representation and mind.* Cambridge, MA: MIT Press.

Poulsen, D., Kintsch, E., Kintsch, W., & Premack, D. (1979). Comparison of young and older children's comprehension of out-of-order picture stories. *J. Exp. Child Psychol., 28*, 379–386.

Premack, D. (1983). The codes of man and beasts. *Behavioral and Brain Sciences, 16*, 246–270.

Premack, D. (1984). Pedagogy and aesthetics as sources of culture. In M. Gazzaniga (Ed.), *Cognitive neuroscience.* New York: Plenum Press.

Premack, D. (1986). *Gavagai! Or the future history of the animal language controversy.* Cambridge, MA: MIT Press.

Premack, D. (1988a). Minds with and without language. In L. Weiskrantz (Ed.), *Thought without language.* Oxford: Clarendon Press.

Premack, D. (1988b). "Does the chimpanzee have a theory of mind?" revisited. In R.W. Byrne & A. Whiten (Eds.), *Machiavellian intelligence.* Oxford: Oxford University Press.

Premack, D. (in preparation). Social cognition.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515–526.

Shultz, T. (1982). Rules of causal attribution. *Monographs of the SRCD, 47*, 1–51.

Spelke, E.S. (1982). Perceptual knowledge of objects in infancy. In J. Mehler, M.F. Garrett, & E.C. Walker (Eds.), *Perspectives in mental representation.* Hillsdale, NJ: Erlbaum.

Rawls, J. (1970). *Theory of justice.* Cambridge, MA: Harvard University Press.

Wellman, H. (in press). *Children's theories of mind.* Cambridge, MA: MIT Press.