Intentional relations and social understanding

John Barresi
Chris Moore
Department of Psychology, Dalhousie University,
Halifax, Nova Scotia B3H 4J1, Canada
Electronic mail: jbarresi@ac.dal.ca; moorec@ac.dal.ca

Abstract: Organisms engage in various activities that are directed at objects, whether real or imagined. Such activities may be termed "intentional relations." We present a four-level framework of social understanding that organizes the ways in which social organisms represent the intentional relations of themselves and other agents. We presuppose that the information available to an organism about its own intentional relations (or first-person information) is qualitatively different from the information available to that organism about other agents' intentional relations (or third-person information). However, through the integration of these two sources of information, it is possible to generate representations of intentional relations that are uniformly applicable to the activities of both self and other. The four levels of the framework differ in the extent to which such integration occurs and in the degree to which imagination is involved in generating these representations. Most animals exist at the lowest level, at which integration of first and third person sources of information does not occur. Of nonhuman species, only great apes exhibit social understanding at intermediate levels, at which integration of these sources of information provides uniform representations of intentional relations. Only humans attain the highest level, at which it is possible to represent intentional relations with mental objects. We propose that with the development of the imagination, children progress through three stages, equivalent to the later three levels of the framework. The abnormalities in social understanding of autistic individuals are hypothesized to result from a failure to develop integrated representations of intentional relations.

Keywords: animal cognition; autism; development; evolution; imitation; intentionality; joint attention; representations; social understanding; theory of mind

1. Introduction

Humans are adapted to living in social groups with complex patterns of social interactions. In recent years, it has been suggested that the evolution of intelligence in primates that ultimately led to human beings was driven in part by the demands of social information processing (e.g., Humphrey 1984). To deal with these demands, the reasoning goes, certain primates, most notably humans, have evolved a conceptual system that is used for making sense of, predicting, and manipulating the behavior of conspecifics (Byrne & Whiten 1988). Adopting a phrase from Premack and Woodruff's (1978) work with chimpanzees, this conceptual system has been called a "theory of mind," and it has become the focus of increasing attention – especially in primate and human developmental research (e.g., Astington et al. 1988; Frye & Moore 1991; Pernier 1991; Wellman 1990; Whiten 1991a).

Whether or not one accepts the label "theory of mind" (for criticisms, see Hobson 1991; Johnson 1988), there are two essential capacities of the human conceptual system for social understanding. First, the conceptual system represents activities of agents that are directed at objects. Such activities include simple purposive actions or psychological orientations that are directed at real objects, such as seeing, fearing, or poking them, as well as more complex mental activities, such as beliefs and desires, which may be directed at imaginary objects. We refer to all such activities involving an agent, a directed activity, and an object as "intentional relations." Second, the conceptual system can be used for understanding the activities of both oneself and others. In other words, we have a single conceptual system that can equally well represent our own activities and the activities of others. Such uniform representation occurs despite the fact that the information available about one's own activities is qualitatively different from the information available about others' activities. It is the relationship between these two capacities of the system that is our main concern in this target article. As we will argue below, the uniform representation of the intentional relations of both self and other requires that the information available about one's own activities be combined with the information available about others' activities. In elaborating our view on how this is possible, we present a framework of forms of representation of intentional relations that identifies four different levels of social understanding. This framework is then applied to evidence from animal phylogeny and human development and to the developmental disorder of autism, which is characterized by a deficit in social understanding.

2. Intentional relations and their representation

2.1. Intentional relations. Organisms engage in a variety of activities. Consider the following examples:
1. The bird builds a nest.
2. The monkey fears the snake.
3. Columbus sees land.
Barresi & Moore: Intentional relations

4. The infant wants food.
5. The frog mistakes a BB for a fly.
6. John thinks that Mary wants a computer for her birthday.
7. Bob remembers that he vacationed in Jamaica last year.
8a. Beatrice knows that Amanda likes Colin.
8b. Beatrice knows that she likes Colin.

In each example an organism directs its activities toward an actual or possible object or state of affairs, and these activities depend on the sensorimotor or higher mental capacities of the organism. We refer to such directed activities as "intentional relations," and all such relations involve three constituents: the agent, the object, and the activity connecting agent to object.1

The examples listed here illustrate a number of features of intentional relations that are important to distinguish. Examples 1–3 present three main types of simple intentional relations. Example 1 involves the action-oriented intentional relation of building a nest and example 2 involves the emotional/motivational intentional relation of fear of a snake. Example 3, involving seeing, is an illustration of epistemic intentional relations, which include sensing, perceiving, knowing, and believing. The difference between examples 1–3 and 4–5 shows that intentional relations can take objects or states of affairs that are current or are in the past or future (or are even just possibilities, including counterfactual ones).

Example 6 illustrates that intentional relations can themselves be the object of other intentional relations. In such cases, the latter intentional relation may be referred to as "second order." In this example, John is the agent of a second order intentional relation for which the object is the first order relation, "Mary wants a computer for her birthday." Example 7 shows that, agents can have second order intentional relations to their own first order intentional relations. Finally, examples 8a and 8b show that a single agent can be in second order intentional relations to comparable first order intentional relations of self and other.

It is second order intentional relations such as 6, 7, and 8 with which we are ultimately concerned in this paper, for it is in these examples that an agent exhibits social understanding. In these particular examples, an agent is in a second order epistemic relation which takes as its object a first order intentional relation involving either the self as agent or another agent. It is under such circumstances that it is usually understood that the agent "represents" or encodes the embedded first order intentional relation of the object-directed activity of an agent (cf. Dennett 1987 for an example of a similar hierarchical system involving, e.g., beliefs about beliefs). Furthermore, and most importantly, it is typically assumed (and explicitly by those who take a "theory of mind" perspective, e.g., Gopnik 1983) that the same representational form is used in encoding the embedded intentional relation, whether the latter involves the self as agent or another agent. For example, in 8a and 8b the same agent (Beatrice) is in second order epistemic relations to two different first order intentional relations. In one of the first order relations, the agent is a different person (Amanda), whereas in the other the agent is the same (Beatrice herself). These examples assume that Beatrice is able to represent the intentional relations of both herself and other agents in the same manner. However, this seemingly innocuous assumption of a common representa-

tional system that is applied to self and others actually has a significant interpretive problem, namely that the information we have about the first order intentional relations of self and other is of qualitatively different kinds. For instance, in examples 8a and 8b, the information that Beatrice has about her own liking of Colin is not the same kind of information that she has about Amanda's comparable liking of Colin. So the problem is: How is it possible that the same representational form can be applied equally to self and other?

2.2. First and third person information about intentional relations. An individual organism's information about intentional relations is available from two different sources. An organism has direct first person information about its own intentional relations and direct third person information about the intentional relations of other agents. In what follows we will argue that such first person information is qualitatively distinct from third person information.2 Before we begin, it is crucial that the reader not presuppose an adult human understanding of intentional relations. There are two illusions associated with adult human knowledge of intentional relations. In the "illusion of first person knowledge" (Gopnik 1983; Moore & Barresi 1983), adult humans believe that they have direct access to their own mental states, such as their beliefs and desires, but that they must infer comparable mental states in others. In addition, there is the equally pervasive "illusion of third person knowledge," whereby adult humans believe that they can perceive directly certain intentional activities of others, such as the latter's seeing an object. In both cases, however, the illusion of direct access may be the product of acquired knowledge of intentional relations. To appreciate the information that is available to simpler organisms, such as animals and young children, it is necessary to dispose of a priori assumptions about the direct accessibility of intentional relations. Instead, it is necessary to consider the information that is directly available to an organism when it attends to the intentional relations of itself or another agent.

When an organism attends to the intentional relations of another agent, the third person information that is immediately available consists most importantly in the behavioral expressions of the other's intentional relations. This information about the other's intentional action is available as gross physical movement; information about the other's emotions is available, in part, as facial transformations; information about the other's perception may be available as orientation of head and eyes. These behavioral expressions may occur in spatiotemporal relations with other objects. For example, an observer attending to another agent who is looking at an object has direct visual access to the latter's head and eye orientation as well as to the spatial array of objects. In cases where the agent is looking at a static object, third person information about the agent's looking may be entirely drawn from the agent's head and eye behavior; there is no direct information available to the observer that the agent's head and eye orientation are directed at an object. In other cases, third person information may also include the spatiotemporal relations between the agent's behavior and the object, such as when the agent visually tracks a moving object, thereby moving its eyes or head in tandem with the object. Even though information about the object may be included in such cases, there is still
no information about the object-directedness of the agent's activity.

Because third person information depends so heavily on spatiotemporal relations involving the agent's movements and these movements' relations to objects, the extent to which such relations are available to the observer varies. For objects that are physically handled or near the agent, it is relatively simple for the observer to form an associative connection between the agent's movements and the object around which the movements are organized. However, for objects at some distance in space, and especially in time, such associative connections are much more complex. So the observation of an agent's head turn may lead the observer to turn in the same direction, thereby allowing the observer to see the object of the other agent's gaze, but the information about the object may not be incorporated as part of a relation between the agent and the object. Although the object of the other agent's gaze is now available to the observer, the latter may not link the object to the agent. In sum, third person information about intentional relations consists entirely in the movements of agents and certain spatiotemporal relations with other objects. Although third person information may be in part information about the organization of the agent's movements in relation to objects, it is not information about the intentional directedness of these relations.

When we consider the information that is available to an agent about its own intentional relations, what is most striking is that this information is qualitatively very different from that which is available about the relations of others. Consider again an agent tracking a moving object. The information available to the agent about its own looking is not visual information about its head and eye movements. The first person information is, rather, predominantly about the object. Nevertheless, there is some first person information about the agent's own activity in relation to the object (Russell 1994). This point is most evident from a comparison of world-caused and self-caused perceptual effects. In terms of visual experience, the apparent movement of an object caused when that object travels in the world from left to right is no different from the apparent movement of an object when the agent moves its eyes from right to left. Yet, organisms are able to distinguish real movements in the world from apparent movements caused by self-motion, so some form of monitoring of the difference must occur. Agents must be able to distinguish perceptual events that are dependent on their own activity from perceptual events that are independent of their own activity. It is not the case that such discrimination requires an understanding of self as agent. Indeed, the above example demonstrates that any motile organism must have this capacity, and it has been shown to exist in relatively primitive organisms such as Drosophila as the phenomenon of different feedback (see Russell 1994). In higher organisms, a variety of forms of monitoring exist, including kinesthetic and proprioceptive feedback. Through these forms of feedback the organism is provided with information about its own actions or possible actions in relation to perceived or imagined properties of the object. The point here is that an agent's experience of its own intentional relations must always involve objects and events in the world in relation to its own activities directed at those objects or events but need not include knowledge of itself as an agent.

In brief, then, first and third person kinds of information about intentional relations that are not already infected by the development of adult human forms of knowledge are qualitatively different.

Nevertheless, adult humans do represent their own intentional relations to objects and those of others in the same manner. It is our view that essential to the development of this capacity is the existence of an "intentional schema." This schema takes as its inputs both first person information about intentional relations and third person information about intentional relations. Because third person information tends to be about agents and their movements in spatiotemporal relations to objects, whereas first person information tends to be about objects in relation to the agent's actions or potential actions, it becomes possible, using the schema, to integrate these two sources of information into a single representation that can be applied equally to the activities of self and other.

2.3. The intentional schema. Schemas are structures that organize information according to rules to yield new perceptual or conceptual units. Whereas some schemas organize information within a single perceptual modality, such as vision or audition, others provide rules for integrating information across modalities. By the intentional schema we mean an intermodal perceptual and conceptual structure with the capacity to coordinate and integrate first and third person sources of information about object-directed activities into representations that link agents to objects through intentional relations. Because first person information about these activities consists primarily of kinesthetic and proprioceptive information, whereas third person information about these activities consists primarily of visual information, the intentional schema is a special case of intermodal schema. Intermodal integration yields representations that preserve the quality of both forms of informational contribution. For example, integration of the visual information about a speaker's mouth movements and auditory information consisting of speech sounds yields a cross-modal representation of speech for which neither form of information is redundant (McGurk & MacDonald 1976). Kinesthetic feedback information may be similarly combined with information from other modalities, for example, when the visual image of oneself in a mirror becomes integrated with one's somatosensory awareness of one's movement so that one can use a mirror to shave or put on makeup. In these two examples, the information that is being integrated originates from a single individual — another person in the speech perception case, the self in the mirror case. What is different about the intentional schema is that informational inputs from both self and other are integrated, and the resultant representation can be applied to either self or other. Consider our previous example of an agent who turns in a certain direction to look at an object; an observer of this agent then turns in the same direction and sees the same object. Although these two events may occur without the observer being aware of their cooccurrence, the observer might also attend to the joint activity of self and other, and integrate the available first and third person information. Through the use of the intentional schema, the observer could generate a representation of the intentional relation that could be applied equally to the looking activity of either self or other. Without the intentional schema, the representation of the looking activity of others would necessarily remain distinct from the representation.
of one's own looking, and neither would resemble the adult human concept of an agent looking at and seeing an object.

We assume that two critical conditions must be met for the intentional schema to generate an integrated representation of a particular intentional relation. First, there must be available to the schema both first and third person information about the intentional relation. If one agent is in a situation with another agent where both agents are engaged in the same intentional relation with an object, then these two sources of information would be presented in current experience. The agent would have first person information about its own intentional relation with the object concurrently with third person information about the other agent's comparable intentional relation with the object. Although there are various means through which this situation can come about, for example, emotional empathy and joint attention, the imitation by one agent of another's actions on objects provides the most general one. Thus, when two individuals exhibit matched behavior toward a shared object while engaged in interaction with each other (as in imitating actions on objects), there is available to each individual both first person kinesthetic and proprioceptive information about its own current intentional relation to the object and third person visual information about the comparable intentional relation of the other. We suppose that it does not matter which individual first generates the matching action as far as the schema is concerned. As we shall see later, an adult might match an infant's action, and thereby provide the infant with its initial opportunity to use the schema by presenting third person information of a comparable action coincident with first person information of the infant's own action.

It is important to emphasize that mere imitation of movements will not suffice to provide an opportunity to generate a representation of an intentional relation. The imitation must be, rather, of the object-directedness of activities. Although imitation per se may not be a necessary condition for the use of the schema, we suggest that it was only with the emergence in certain species of a capacity to imitate the object-directed activity of others that the appropriate general conditions occurred for the use of the schema. This mimetic capacity provides a way to generalize the matching of actions between self and other and allows the organism to generate first person information of one's own actions to match with a wide range of third person information obtained from observing the actions of others. Indeed, as we shall soon show, the evolutionary emergence of such an object-directed imitative responsivity— at least among primates— seems to have occurred in just those species for which we have independent evidence of a type of social understanding that suggests the use of the intentional schema.

We also suggest that in cognitively advanced organisms not all uses of the schema require that both observer and actor currently engage in matched actions, but that one or both of the sources of information may be derived not from perception but from the imagination. By "imagination," we mean the capacity to mentally represent or think about objects, in their absence, by using memory-based information.

Through imagination, noncurrent first or third person information can act as input for the intentional schema. For example, one individual may observe another's behavior directly (third person information) and imagine the first person information associated with the object or situation at which the action is directed. Conversely, one individual may engage in an action directed at an object or situation and thereby directly obtain the first person information of that action and, at the same time, that individual may imagine how such an action may appear to an observer, thereby imagining the third person information appropriate to that action. Such imaginative acts will also provide matched first and third person information, which can then be integrated through the intentional schema to generate a representation of the intentional relation. The resultant representation unifies these two sources of information in a form that would apply equally to self or other regardless of the original source of the current nonimagined information.

The second condition for the generation of an integrated representation of an intentional relation by the intentional schema is that, when the first condition of matched first and third person informational input is met (either in actuality or in imagination), the observer is able to attend to both informational elements. For example, in the case of imitation, regardless of who first generates the matching response, one individual must attend to the other's behavior, while at the same time attending to the first person information about its own comparable intentional relation. Attention to both informational sources provides both inputs, so that a representation of the intentional relation may be generated through the intentional schema.

It is only when both of these conditions are satisfied, that is, (1) matched first and third person inputs and (2) attention to these inputs, that the intentional schema can be used to integrate first and third person information into a single representation of an intentional relation. Whether the intentional schema should be considered a special module evolved for the particular purpose of social understanding or whether general purpose information processing mechanisms are involved is not our particular concern at this point. That is an empirical issue. What is essential is that some such mechanism is in place for some species (most obviously humans) and not for others, even though the latter may satisfy the two conditions that have been considered. Moreover, this mechanism must join first and third person information of agents' relations to objects in a particular way so that the resultant representation can be applied in a comparable manner to self and other, thus facilitating the kinds of social understanding found in human beings and some other primates.

2.4. A four-level framework of representations of intentional relations. In this section we present a framework for the forms that representations of intentional relations take. This framework identifies four levels of representation of intentional relations (see Table 1). It is derived in part from an empirical base, which will be discussed in detail in later sections. At the lowest level (level 1) are representations that treat first person information about the organism's own intentional relations differently from third person information about the intentional relations of others. Because, as we have seen, first person information has little in common with third person information, knowledge about the intentional relations of self gained through first person information is quite distinct from knowledge about the intentional relations of others based on third person information. At this level, all representation of intentional relations treats first and third person information about intentional rela-
<table>
<thead>
<tr>
<th>Level</th>
<th>Information input</th>
<th>Representations</th>
<th>Phylogeny</th>
<th>Human ontogeny</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1st person</td>
<td>1st person perspective (self)</td>
<td>Animals with expectancies (e.g., rats, monkeys)</td>
<td>Autistic</td>
</tr>
<tr>
<td></td>
<td>3rd person</td>
<td>3rd person perspective (other)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1st person current</td>
<td>Shared intentional relation to object (self and other)</td>
<td>Certain animals in narrow range of functional contexts (e.g., vervets)?</td>
<td>9-month-old</td>
</tr>
<tr>
<td></td>
<td>3rd person current</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1st person current</td>
<td>Concept of intentional agent (self or other)</td>
<td>Great apes, Cetaceans?</td>
<td>18-month-old</td>
</tr>
<tr>
<td></td>
<td>3rd person current</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>or</td>
<td>1st person imagined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3rd person current</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1st person imagined</td>
<td>Concept of mental agent (self or other)</td>
<td>Humans</td>
<td>4-year-old</td>
</tr>
<tr>
<td></td>
<td>3rd person imagined</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Four levels of social understanding indicating informational inputs and representations generated.

As a result, there are two distinct types of knowledge about intentional relations, one that applies only to other individuals and one that applies only to self.

Given the capacity to integrate first and third person information through the intentional schema, we see three additional levels in the form of such representations. The two kinds of information may be derived from direct or current experience or from noncurrent, imaginative sources. The three levels at which the schema is used differ in the degree to which imagination is involved in the generation of the first and third person information that is integrated.

At level 2, both inputs are provided through current, immediately available sensory and perceptual information as opposed to noncurrent, memory-based, imaginative information. An organism can be said to be operating at level 2 when it is capable of attending to and integrating the first and third person sources of information available when a particular intentional relation to an object is shared with another individual. At this level, the similarity between self and other need not be recognized, but the organism must show evidence of knowledge of the shared relation in the sense that its actions will differ depending upon whether the relation is shared or not. In some circumstances the organism may attempt to establish a shared relation with a conspecific. For example, 12-month-old human infants, while not showing a clear awareness of themselves as agents, will make communicative attempts to engage others in shared attention to an object. At level 2, then, intentional relations are not understood to be associated with a single individual, either self or other. Nevertheless the representations of activities do necessarily involve both first and third person information and thereby capture the intentional nature of the relation between an agent and object.

At level 3, current experience of an intentional relation provides one input to the schema and imagination provides the second. The current input, whether from self or other, instead of leading to an attempt to generate matched input through an action, leads to the generation of this input through an act of the imagination based on memorial representations of similar intentional relations. The intentional schema is then used to yield a new representation of an intentional relation based on only one perceptual input. The presence of imaginal input to the intentional schema means that an integrated representation of a particular intentional relation can be generated without a matching of action. As a result, at this level it is possible to represent the current intentional relation of an agent, whether it be self or other, and distinguish it from the intentional relations of other agents. Thus the self–other distinction can become explicit, and the individual can be said to have a concept of an intentional agent. The understanding of the intentional relations of both self and other, however, is still tied to current experience. In the case of the other, the third person information is typically given by current perceptual input and the first person information is imagined. For example, an individual is observed to display a particular facial expression in the direction of an object and the observer imagines the first person component of the emotional attitude corresponding to that display. In the case of the self, typically, it is the third person information that is imagined, while the first person information is provided by the intentional relation in which the actor is immediately engaged. Feeling an emotional orientation to an object thus becomes integrated with an image of how that orientation would appear to an observer. The understanding of intentional relations that results at this level is one involving diversity in the intentional orientations of different agents to a common directly accessible situation (though it may be a "pretend" situation). Because direct access to a current situation constrains this form of intentional representation, it cannot represent diversity in intentional relations of agents to different mental objects – some of which may misrepresent the current situation.

Finally, at level 4, both inputs to the schema come from the imagination rather than through current experience. Situations necessitating level 4 are those that require the representation of an intentional relation of either self or other where neither current first nor third person information enters into the intentional relation. In order to resolve this problem, the intentional relation, either of self or other, must be abstractedly represented with an agent's orientation to a mental object. In these cases, as with level 3, current experience typically activates stored representations of intentional relations and it is these representations that provide the inputs for the schema. In contrast to level 3, however, the novel schematic representation is now entirely composed in the imagination, in the sense that both first and third person information are imagined and integrated through the schema. Thus, the observer can imagine the third person information of an agent in some situation.
and, at the same time, can imagine that agent's first person orientation to that situation. In this way, the agent is placed in a situation different from that which the observer faces, and the novel representation of the imagined intentional relation can be compared to the current experience of the observer. As a result, the mental object in the imagined situation is cleanly distinguished from the actual object in the situation as perceived by the observer. The individual at this level, then, may be said to have a concept of a mental agent. Although there may be further levels, such as that of a reflective agent (see Barresi & Moore 1993; Tomasello et al. 1993a), our account of the understanding of intentional relations—particularly in human development—terminates at this level of representation, since it is at this level that misrepresentation can be understood by an agent.

Now that we have a framework indicating various levels of the representation of intentional relations, we can use it to describe how organisms come to understand themselves and others as intentional beings. In the next section, we suggest that most animals operate at the lowest level of the framework. Among nonhuman animals, only great apes appear to exhibit a general social understanding beyond level 1, though they do not attain level 4. In the following section, we present evidence that during the first four years of life, children progress through a series of three stages, equivalent to the later three levels of the framework. Finally, we argue that, perhaps as a result of a failure in information integration, autistic individuals fail to develop representations of intentional relations that can be applied uniformly to self and other.

3. The phylogensis of social understanding

Nonhuman species can be divided into two classes based on whether or not they can categorize their own and others' intentional relations using a single conceptual system. Because most organisms lack this special capacity, they tend to interpret their own and others' actions using different conceptual systems. For these organisms, most representations based on first person information are distinct from those based on third person information, and there is no general understanding of the similarity in intentional relations between self and other. We suggest that these organisms operate for the most part at level 1 of our framework and sometimes possibly at level 2, typically in communicative domains, where they may recognize the comparability of their own activities and those of others (see Table 1). Members of a smaller class of organisms are able to interpret the activities of themselves and others using the same conceptual system. The evidence that this capacity is present is the ability of some species, in particular the great apes, to imitate novel goal-directed actions. These species can operate at levels 2 and 3 but not at level 4 of our framework (see Table 1). We do not think that this mimetic capacity alone is sufficient to allow them to appreciate the representative character of mental activity. Only humans have the capacity to think about the representational nature of mental relations and thus to achieve level 4 of the present framework.

3.1. Level 1 or nonimitative organisms. Insofar as a level 1 organism is unable to apply the same conceptual system to self and other, such an organism will be severely restricted in what it can appreciate of the meaning of its conspecifics' activity and what it can learn from that activity. Although we do not wish to restrict ourselves to particular models of learning, it is worth considering the nature of level 1 processing with respect to Pavlovian and instrumental conditioning from the perspective of the expectancy theory of Dickinson (1989; Heyes & Dickinson 1990). For level 1 organisms, learning about the intentional relations of others takes the form of Pavlovian, stimulus–stimulus generalizations or expectancies about one stimulus given another. Learning about their own intentional relations takes the form of instrumental expectancies or expectancies about outcomes given particular actions. In observing others, one organism can form stimulus–stimulus generalizations and expectancies between different observable behaviors of the others, for example, between a grimace and a likely response of another organism in a certain context. It can also notice stimulus–stimulus relations between observable behaviors of the organism and objects or events in the world, for example, the orientation of the other organism's head or eyes in a direction where an interesting object might be found. In this way, much can be learned about the intentional relations of others, both general and specific. It may be possible to learn to detect deceptive acts by others from the absence of the usual contingency between an observed act and its consequence. An organism will neither be able to interpret these stimulus–stimulus complexes in terms of the motives of the other organism, however, nor to see the other's movements as actions that are analogous to its own. Though it may act in response to the deceptive behavior, it will not be able to recognize that it is responding to the behavior as deceptive.

In contrast to the stimulus–stimulus learning that can occur through the observation of others, a level 1 organism will be able to acquire knowledge of its own instrumental actions and their expected consequences. Many organisms are capable of forming expectancies that represent the relation between their own instrumental behavior and a consequence (see Dickinson 1989). Unlike Pavlovian conditioning, which does not require the representation of one's own potential behavior, instrumental conditioning has been interpreted as involving a representation of the action in its relation to the expected consequence or goal. Thus, for instance, if the reward that is "expected" to occur following a particular response is decreased, rats but not goldfish will begin to exhibit an instrumental response depressed below the control level for the new reward. This result has been interpreted to suggest that rats have representational knowledge of the behavior–reward contingencies, but goldfish do not (Bitterman 1987; Dickinson 1989; Mackintosh 1987, see Macphail 1987 for a contrary view). Assuming that this is the case for instrumental learning for some organisms, an organism with this capacity, but still at level 1 of our framework, would have a representation of its own goal-directed action in a manner that it could not generalize to the comparable goal-directed actions of others. Nor could it translate the goal-directed actions of others into actions it could adopt for itself. A result of this lack of comparability between the actions of self and others is that such an organism could learn to engage in fairly elaborate deceptive behaviors without recognizing that its behavior was, in fact, deception; at the same time it could be easily deceived by a comparable deceptive act of the other.
In short, the activities of self and other are processed in different ways by level 1 organisms. The visual representation of the activity of others that may act as a conditioned stimulus is not directly translatable into the representation of the organism's own instrumental activities. We have suggested that the advent of the ability to interpret the activities of both self and other within a single conceptual scheme comes with a generalized capacity to imitate goal-directed activity. Such a capacity demonstrates crossmodal matching not only of body movements based on visual and kinesthetic information (Mitchell 1994) but also of those directed at objects or goals. It is the latter forms of movement that must be matched for uniform processing of the intentional relations of self and other.

In evaluating our framework with respect to lower organisms it is important to distinguish similarity in the behavioral manifestations of intentional relations caused by a general capacity to imitate goal-directed actions and such similarity caused by alternative mechanisms. Among many – if not all – social species, a wide variety of mechanisms produce similarity in behavior in the absence of a general capacity to imitate novel behaviors, whether expressive or goal-directed. In a recent review of the literature, B. Moore (1992) has described 10 mechanisms by which 5 organisms might exhibit similar behavior in a particular context without having the capacity to imitate novel actions (see also Gallef 1988; Whiten & Ham 1992). In what follows we will give a brief overview of some of the matched activities we believe can be accounted for at level 1 (or at most level 2) of our framework. We will then go on to consider more complex phenomena by nonimitative organisms that might be interpreted as higher levels of our framework, but we will suggest how they might be accounted for at level 1. Finally, we will consider recent experimental studies consistent with this view, which suggest that monkeys do not exhibit the generalized capacity to understand the viewpoint of others indicative of level 3 processing.

A simple form of such sharing of intentional relations occurs when an organism is a member of a group and responds to the perceived third person aspects of the intentional relations of others by adopting the same intentional relation itself. Examples of such sharing of action relations can be seen in the herding behavior of mammals and schooling behavior of fish. Examples of sharing of expressive behavior that serve to communicate emotional/motivational intentions are the croaking of frogs, the barking of dogs, and the howling of wolves. There are also examples of the sharing of epistemic relations in animal systems. Upon seeing a conspecific stand to look in a particular direction, baboons will themselves adopt a similar orientation; this mechanism thus allows each baboon to benefit from the earliest exposure of any member of the troop to a significant environmental event. Such shared intentional relations come to play a quite precise role in more complex forms of communication of emotional as well as epistemic relations. Cheney and Seyfarth (1990) have described how vervet monkeys use particular calls that not only express an emotion, such as fear, but indicate the presence of particular predators. Such signals elicit appropriate evasive action in a heater and also cause it to generate the same call, thus producing a unified group activity of communication and action.

We believe that most of these phenomena can be accounted for at level 1 of our framework without proposing the existence of an intentional schema that applies uniform representations to the actions of self and other. In these cases, we would suggest that each group member has only a third person representation of any other's activity and a first person representation of its own activity. However, the other's acts, such as movements or emotional expressions, come to be associated with the first person aspects of those expressions as experienced concurrently or in close synchrony. Because of the close synchrony of third and first person representations of common intentional relations, the individual organism does not seem to act as an individual but as a member of a group. Such associated, but not integrated, first and third person representations of intentional relations act nearly as well for group activity as an integrated version at the next level of our framework. For such complex communicative systems as the vervet calls, it seems possible that the organisms recognize the similarity of calls being used by self and other (e.g., the "moving into the open" grunt studied by Cheney & Seyfarth, 1990; 1992; also described by Dennett, 1987). If so, then these organisms are operating at level 2 of our framework; such cases, however, seem to be the exception rather than the rule.

With the possible exception of complex communication systems, we believe that the majority of the behaviors of self and other are not represented in the same representational system by nonimitative organisms. To further support our claim that even monkeys' social skills can be accounted for at level 1, it is worth evaluating some recent research on the ability of monkeys to attribute intentional relations to conspecifics. Cheney and Seyfarth (1990) conveniently provide a review of this research, particularly in their Chapter 8, on attribution. Their summary indicates that monkeys are capable of both interpreting current stimulation from conspecifics that might have consequences for their immediate goal attainment and anticipating the responses of conspecifics to their own actions and expressive behavior. Insofar as intentional relations of others can be understood based on perceived properties of objects and the contingent relations of other monkeys to those objects, or, in other words, from a third person perspective, these monkeys seem highly skilled in anticipating the actions of others based on these relations. However, monkeys seem incapable of interpreting the behavioral tendencies of others as a reflection of the observed monkey's motives, emotions, and epistemic relations in situations where the observers are -- in a sense -- disinterested. They exhibit virtually no signs of empathy, defined as the capacity to recognize the emotional relations of others that they themselves are not experiencing. For example, they do not exhibit compassion by comforting another monkey in grief over the loss of a close relative or by providing food to the old and disabled.

Self-knowledge of monkeys is likewise limited and of a form different from their knowledge of others. For instance, though they can learn to use mirrors to perceive the activities of others, they do not recognize themselves in mirrors. That their knowledge of their own intentional relations is represented differently from their knowledge of the intentional relations of others is indicated by a number of the findings. For instance, they have sufficient awareness of their own goals to avoid being deterred from achieving them when they anticipate that another will interfere. Consistent with expectancy theory (Dickinson 1989), the agent's own goals and knowledge of how to achieve them.
are not immediately visible to the other, yet do determine its own actions.

The degree of sophistication that monkeys can achieve while remaining at level 1 of our framework is probably best exhibited in the various forms of tactical deception found in baboons. Whiten cites an example provided by Strum (1991b). A male baboon, Dr. Bob, was tricked by another baboon, Peggy, who groomed him until he released his grip on an antelope that he had refused to share, so that she could grab it and run. When a similar situation arose later, Dr. Bob held on tightly despite the grooming. Peggy, not to be thwarted, then attacked a female associate of Dr. Bob's until he — apparently with trepidation — abandoned the meat and came to her defense; Peggy again grabbed the meat and ran off. Now, it is easy enough to imagine that Dr. Bob and Peggy, especially, understood the intentional relations of each other and themselves in the same conceptual terms. But this is by no means necessary in the present case. Both Dr. Bob and Peggy could perceive the other's actions and potential actions entirely from a third person perspective and their own goals and possible actions from a first person perspective, without in the least being able to imagine how their own actions appeared to the other or how the other's actions appeared to that other. In both directions, the appropriate actions to take and the likely actions of the other could be based on the immediate perception of the situation and an acquired expectancy of the relation between one's own potential action and the behavior likely to be exhibited by the other organism in response to it. Hence, without further evidence that the organism has knowledge of the perspective of the other animal that could not have been acquired through instrumental learning or stimulus-stimulus generalizations, even such sophisticated acts of deception are easily handled by representations acquired at level 1 of the present framework.

There are two kinds of phenomena that would indicate that monkeys really understood their own and the other's behavior from the perspective of the other animal. First, such understanding would be shown by the imitation of novel purposive actions of conspecifics based purely on the visual observation of the performance of those actions. Second, it would be shown by the ability to switch roles with another organism with whom they have jointly learned a cooperative task. Yet, the evidence to date indicates that monkeys show neither of these abilities.

Evidence of the monkey's ability to imitate novel purposive action is quite limited (Cheney & Seyfarth 1990; Galef 1990; Whiten & Ham 1992). For instance, consider the famous case of the potato washing technique acquired through a kind of cultural transmission among Japanese macaques. What is notable, besides the extremely slow diffusion of this skill, is that the monkeys seemed to acquire only knowledge of third person contingent relations involving stimulus complexes of water, edible potatoes, and monkey movements with these objects, not first person knowledge of how to wash potatoes or direct awareness of the function or goal of washing. As a result each monkey independently had to make sense of the stimulus complex to which it was exposed, reinventing the act of washing the potatoes and thus developing its own peculiar method. If the monkey's perception of the actions of others represented those actions in the same way as its own actions, as we suggest occurs at level 2 or higher in our framework, then we believe that imitation would have presented the "how" and "why" as well as the "what" of potato washing.

Further evidence that monkeys are probably not able to represent their own and the other's actions in a common representational form is provided by a recent study by Povinelli, Parks, and Novak (1992). They constructed a paradigm in which a rhesus monkey cooperated with a human, each performing different actions in order to generate a reward for the monkey. Different monkeys were trained on each of the two possible roles. After having acquired the ability to play one of the two roles, the monkeys were assigned to the opposite role. If the monkeys had represented the actions of both agents within a common representational system, then the role reversal should have produced strong positive transfer, as it did in a previous study with chimpanzees (Povinelli et al. 1992). None of the monkeys, however, showed immediate comprehension of their new role. This suggests that the monkeys' representation of their own role was different from their representation of the role of the other agent and that this is because monkeys — unlike chimpanzees — are restricted to level 1 representations. Hence the representation of their own actions uses only first person information, whereas the representation of the actions of the other agent uses only third person information and there is no direct translation between these representational forms that could be used in the role transfer task of the experiment.

3.2. Imitative organisms. Cheney and Seyfarth (1990) suggest that only higher primates have a general ability to imitate a functionally useful technique. B. Moore (1992), in his review of the imitation literature, also suggests that the evidence of imitation in monkeys is too limited to attribute this capacity to them, and of nonhuman species, it is only the great apes, cetaceans (e.g., dolphins) and certain avian species, in particular parrots, that have this capacity. For our purposes, it is interesting to note that only the most intelligent and social mammals are capable of a generalized imitative ability. We would go further and suggest that the imitative ability of avian species, which functions predominantly for display, is intrinsically different from that of the social mammals. While avian imitation involves only the form of the movement or sound, mammalian imitation is centered around the goal-directedness of activities and appears to function predominantly for social learning. Hence we believe that the general capacity to imitate actions directed at objects for a purpose may be an important skill showing an ability to interpret intentional relations of self and other within a common conceptual system.

In what follows, we will evaluate the capacities for social understanding in great apes with a particular focus on chimpanzees (for gorillas, see Patterson & Linden 1981; for orangutans, see Miles 1986; 1990; Russon & Galikias 1993; for a more general survey that includes dolphins, see Mitchell 1994) and argue that the capacities that they exhibit indicate that they can operate at levels 2 and 3 of our framework as well as at level 1. We would suggest that these capacities imply that they have acquired a concept of an intentional agent (see Barresi & Moore 1993; Tomasello et al. 1993), which they can apply to self and other.

Several interrelated phenomena indicate that adult chimpanzees have the concept of an intentional agent and make use of the intentional schema that joins first and third
person information of an agent engaged in a current inten-
tional relation. Consistent with this hypothesis, chim-
panzees are able to determine the goals of others’ actions
when they observe their movements directed at particu-
lar objects, for example, reaching for a distant object (Pre-
mack & Woodruff 1978). They can also imitate the actions of
others based on the same observations because the rep-
resentation of the action is in a format that can be ap-
plied to the self as well as to the other. Hence, they are
able to imitate both the means and the ends of the actions
(Tomasello et al. 1993b). Since they appreciate the other’s
point of view as well as their own relative to a co-
operative task, they can reverse roles in the Povinelli task and
spontaneously adopt the behavior of the other agent
(Povinelli et al. 1992). Furthermore, chimpanzees have also
been known to show emotional sensitivity and compassion
(Cheney & Seyfarth 1990; Goodall 1990), indicating that
their observation of a conspecific’s expressive behavior is
sufficient to provide an understanding of the other’s emo-
tional intentional relations directed at goals. Chimpanzees
have been reported to show quite sophisticated acts of
deception and counterdeception (Whiten 1993; Whiten &
Byrne 1988). The latter acts seem to require the capacity to
distinguish the “real” goal of the action of self or other from
the “apparent” goal of the behavior, and to recognize that
the apparent goal-oriented act can function to achieve the
real goal. Although not yet requiring a notion of misrep-
resentation, such deceptive behavior depends on the concept
of an intentional agent (cf. Whiten 1993). Finally, although
the interpretation of mirror self-recognition remains uncer-
tain (Heyes 1994; Mitchell 1993), the fact that chimpanzees
but not monkeys do show such self-recognition suggests
that the former are able to interpret their own image in
the same manner as they interpret the images of others, which
would follow if they acquired a concept of intentional agent
that they could apply to self as well as other (Gallup 1977;
1982). Premack and Woodruff (1978; Woodruff & Premack
1979) originally suggested that chimpanzees have a “theory
of mind.” Gallup (1982) went further, hypothesizing that
any organisms capable of self-recognition would also have
an introspective awareness of their own mental states and
the ability to ascribe mental states to others. In a recent
article, Povinelli (1993) appears sympathetic to Gallup’s
interpretation. These suggestions imply that chimpanzees
are able to understand their own and other individuals’
behavior not only at levels 2 and 3 of our framework, but at
level 4 as well. If Gallup’s model is correct, then chim-
panzees would not only be aware of themselves as inten-
tional agents of the same kind as their peers, but they would
also have introspective knowledge of their mental states
and be able to attribute such states to others. We believe
that Gallup’s model of chimpanzee social understanding is
too rich given the current state of the evidence (see also
Heyes 1993). Some of the evidence from Povinelli’s studies
and those of others (see Cheney & Seyfarth 1990) suggests
that chimpanzees have the ability to represent in some
fashion the intentional relations of others, in particular to
represent the motives, emotions, and epistemic relations as
perceived in the others’ actions. They may also be capable
of representing diversity in the knowledge relations of
others, for example, by differentiating an actor who has
knowledge of information from one who does not (Povinelli
et al. 1990).

There is little evidence, however, to suggest that chim-
panzees can appreciate the representational nature of men-
tal activities, and therefore misrepresentations, which re-
quire level 4 representations. Premack’s original work has
been contested (see, e.g., Bennett 1978; Dennet 1978) and,
indeed, led to the interesting developmental studies on
human “theory of mind,” which we discuss later. Pre-
mack’s own subsequent research (e.g., Premack & Dasser
1991) suggests that the chimpanzee, at its outer limit,
behaves like a three-year-old child in its capacities, a level
congruent with level 3 of our framework as we will show in
the next section. An attempt to assess the chimpanzee’s
ability to attribute false belief failed to provide positive evidence
(Premack 1988). Given the limited evidence of knowledge
of intentional relations beyond those expressed in immedi-
ate action, and the absence of positive evidence in favor of
chimpanzees’ and other great apes’ ability to handle ab-
stract epistemic relations, such as false belief, we are not
convinced that the intentional ascriptions of these species
are represented beyond level 3 of our framework. Hence
neither the assumption of a theory of mind nor that of
introspective awareness seem warranted.

When we turn to human social understanding, we find
evidence for all four levels of our framework. In particular,
humans are capable of abstractly representing intentional
relations of self and other, characteristic of level 4. How-
ever, these level 4 representations of intentional relations
should not be considered to be replacements for lower level
representations. It is likely that humans exhibit social un-
derstanding at all levels of the framework depending on the
demands of the circumstances. Furthermore, the human
capacity for level 4 representations of intentional relations
does not appear full-blown at birth. We turn now to the
ontogenesis of human social understanding and apply the
framework to a variety of the phenomena seen in the study
of human development.

4. The development of social understanding

In what follows, we give a schematic account of the de-
velopment of social understanding during infancy and the
preschool years, using a three-stage ontogenetic model
based on levels 2–4 of the framework introduced earlier
(see Table 1). Let us summarize the main points of our
account here. First, we will briefly review the social abilities
of the infant during the first year of life, which, from our
point of view, indicate the use of the intentional schema.
These behaviors, such as joint attention and social referenc-
ing, are ones in which a particular intentional relation is
shared with another; at this stage, the infant can, therefore,
be said to be operating at level 2 of our framework. Once
the first representation of intentional relations is in place,
we suggest that subsequent developments in the represen-
tation of intentional relations depend on domain-general
changes in imaginative ability. These changes may well be
the result of maturational changes in information-processing
capacity (cf. Olson 1989; 1993). During the second half of
the second year, the child becomes capable of imagining
one informational component of an intentional relation
while attending to the other. This ability marks the emer-
gence of level 3 of the framework, where intentional rela-
tions are ascribed to individual agents, including self and
other. The child can therefore be said to have a concept of

BEHAVIORAL AND BRAIN SCIENCES (1996) 19:1 115
information integration is seen clearly in cases of intermodal perception of temporal properties such as rhythm (e.g., Spelke 1979). With respect to social information, Walker (1982) has shown that by 3 to 7 months infants are able to detect the correspondence between visual and auditory information specifying the expression of emotions. However, in this case both pieces of information were of other people and were of the third party variety. Can infants integrate information of the third person sort with information of the first person sort, when those pieces of information are presented via different modalities? Bahrick and Watson (1985) have shown that 5-month-old infants are able to detect the relationship between their own leg movement and a real-time video display of that movement, even when they cannot simultaneously see their leg. They suggest that the basis for this detection is contingency. These results show that intermodal integration of information from vision and proprioception is occurring at this stage in infancy, and it seems quite likely that similar integration could occur for information provided by different individuals if there was a high degree of contingency in the presentation of this information. It could be, for example, that the strong tendency of mothers and others to imitate infants’ behavior in the first year (e.g., Moran et al. 1987) provides the necessary contingent experience for infants to recognize the match between their own and others’ movements.

Recently, Meltzoff (1990; Meltzoff & Gopnik 1993) has gone further and has proposed that innate knowledge of the equivalence of first and third person information has an important role in the development of social understanding. He suggests that “infants, even newborns, are capable of approaching the equivalence between body transformations they see and ones they feel themselves perform” (Meltzoff 1990, p. 160) through the existence of an innate intermodal representational system that codes the body movements of both self and other. The evidential basis for Meltzoff’s claim for such an innate intermodal representational system is the apparent neonatal imitation of facial gestures such as tongue protrusion and mouth opening (Meltzoff & Moore 1977; 1983). However, a recent review and meta-analysis of the research on neonatal imitation by Anisfeld (1991) has shown that reliable evidence has been obtained for only one facial gesture, namely tongue protrusion. Without strong evidence for the imitation of a range of facial expressions, there is no need to postulate the existence of an innate intermodal representational system that would allow the recognition of the equivalence between others’ movements and one’s own.

In any case, the issue of whether the understanding of the equivalence of movements of self and other is innate or acquired from experience is not crucial to the present discussion. Whether or not Meltzoff is right in proposing the existence of an innate representational system for coding invariance in the movements of self and other, and whether or not infants learn to imitate movements, neither case helps generate an understanding of the object-directedness of such movements, because no object is included in the representation. The important consideration is that during the second half of the first year, when objects first start to be brought into interactions between infants and others, infants are able to match the behavior of others. In this way, object-directed activities can be imitated. Once the infant can match others’ behavior toward
objects, they will have available the kind of experience necessary for the first representation of intentional relations (see also Moore & Barresi 1993). It seems plausible that the first intentional relations to be represented are those involving body parts in contact with objects. For example, the infant who is participating in an imitative game involving a particular action on an object will have available concurrently both first and third person information about that action intentional relation. These two types of information may then be combined through intermodal integration into a representation of that intentional relation.

The representation of intentional relations, however, will not be limited to cases where the agent makes contact with the object. Although the infant’s first tendency to turn to look in the same direction as another or to use another’s emotional expression in social referencing may have emerged as a result of learning an innate orienting response (Campos 1983; Gewirtz & Pelaez-Nogueras 1992; Moore & Corkum 1994), we suggest that the infant’s representation of these events is not limited to that seen in lower primates. Once acquired, these behaviors will provide the infant with both first and third person information about the shared intentional relation. Thus, in joint attention, infants observe the head/eye orientation of the other and experience the kinesthetic feedback from their own head turning and the sight of the object. Here, intermodal integration will combine these two types of information into a representation of the intentional relation between viewer and object. In the case of social referencing, the infant observes another’s emotional facial expression and adopts a corresponding emotional orientation to the object of interest. In this case, intermodal integration will combine the first and third person information into a representation of the emotional relation between agent and object. In all these cases, then, the infant’s participation in shared intentional relations with others allows the intentional relation to be represented by integrating the third person information available from the observation of others’ behavior and the first person information available from actually being in an intentional relation with an object. We suggest that the evidence for this integration is seen most clearly in the fact that, unlike the majority of cases of apparent shared intentional activity in other species, infants will check back between observation of the object and observation of the adult (cf. Gomez 1991), thereby indicating concurrent attention to both informational elements.

In sum we are proposing here that, during the first year of life, infants become able to integrate information of the first and third person sort. Toward the end of the first year, the inclusion of objects into social interaction provides the first understanding of intentional relations. Initially, infants understand these intentional relations to the extent that they can participate in episodes of shared relations, with the first person component corresponding to the information available from their own intentional relation to some object or state of affairs, and the third person component constituted by the information available to them from the other’s relation to that same object or state of affairs. At this stage there is a lack of differentiation of the sources of the first and third person information that enters into the experience of an intentional relation, so the intentional relation is not attributable to individual agents. This state of affairs occurs because only immediately experienced first person and third person information enters into the representation of the intentional relation at this time. This situation gradually gives way during the second year of life to one in which self and other are differentiated, with each understood to have both first and third person aspects.

4.2 Stage 2 (level 3): Imagination and the self—other distinction. The second year of life sees the development of a variety of phenomena, persuading many authors that major changes in cognitive abilities are occurring. The development of flexible manual search for hidden objects (Piaget 1953), deferred imitation (Piaget 1962), language use (Bates 1979; Greenfield & Smith 1976), and pretend play (Leslie 1987; Piaget 1962) have all been interpreted in terms of a general change in representational capacity. During the first 15 months of life, infants show an increasingly sophisticated ability to interpret and respond to perceptual events in terms of stored representations and can show recognition and discrimination. However, what is new about the phenomena appearing during the second half of the second year is that not only can infants respond to some perceptual object or event, they can also simultaneously “hold in mind” a representation of a nonpresent or noncurrent object or event (e.g., Olson 1989, 1993). The immense advantage of this representation is that infants are now able to hold in mind objects and events from other places that occurred in the past (as in memory), as well as those that have not yet occurred, but could (Ferner 1991). Furthermore, the mental representation can control behavior, so that activity can now be conducted in relation to imaginary objects or states of affairs rather than real ones, as seen, for example, in pretend play. Rather than “representation,” we prefer to use the term “imagination” to refer to this capacity, because the latter term more adequately captures the idea that this cognitive advance allows the mind to range over both past and future as well as other possible objects and events.

The development of the imagination makes a new stage of social understanding possible. In stage 1 of social understanding, infants understand intentional relations to the extent that they can participate in episodes of shared intentional relations with others, such as joint attention and imitation. Because the understanding of the intentional relation involves both first person information about the intentional relation of the self and third person information about the intentional relation of the other, it cannot be assigned independently to either the self or the other. In contrast, in stage 2, the development of the imagination makes possible intentional relations not given in current experience—third person for self, first person for other. In this way, upon seeing another’s behavior, the child will be able to imagine the first person information of the other. Similarly, while experiencing first person information about its own intentional relation, the child will be able to imagine the corresponding third person information. With both informational components available at the same time for either self or other, the child can integrate them so that a full intentional relation is represented for either self or other. In short, during stage 2 of social understanding, children can be said to have a concept of intentional agency (Barresi & Moore 1993) in the sense that they understand that self and other are separate but similar, both being objective entities that exhibit behaviors directed at objects with accompanying first-person experience of those ob-
jects. In this context, it is noteworthy that children start to use personal pronouns and gestures such as pointing to pick out individual others or the self at about 18 months of age (Bates 1990).

As already mentioned, there are two complementary sides to the understanding of intentional relations in stage 2. On the one hand, the child becomes aware of third person information about the self, which is integrated with first person information about the self. On the other hand, the child becomes aware that others have first person information about their activities in relation to objects, which gets integrated with the third person information about them. Perhaps the clearest sign that the child is aware of third person information about the self is self-recognition. Self-recognition is usually tested using the marked face paradigms developed by Callup (1970) with chimpanzees and Amsterdam (1972) with infants, in which the subject is surreptitiously marked with rouge on the nose and then observed in front of a mirror. Behavior directed towards the subject's own nose as opposed to the mirror in such a situation appears during the second half of the second year (e.g., Amsterdam 1972; Lewis & Brooks-Gunn 1979). Self-recognition indicates that the child is able to imagine self as an object like other objects and in particular like other persons whose third person information the child directly perceives. With the development of the imagination, such imagined third person information of the self becomes a possibility. Strictly speaking, such self-recognition can be achieved without the use of intermodal integration through the intentional schema, because it does not require the imagined third person information of self to be integrated with the first person information of the self (cf. Mitchell 1993). However, at about the same time, children also show evidence of self-conscious emotions such as coyness and embarrassment (Lewis et al. 1989). The latter forms of behavior suggest that children's intentional relations are affected by their imagination of how they appear from the intentional perspective of another. If this interpretation is appropriate, then these emotions would require not only imagined third person information of self but an integration of first with third person information of the self through the intentional schema.

To determine unambiguously that children understand that others have first person information, it is necessary to show that they can behave appropriately in relation to the other's first person information rather than their own. Again, the first clear signs of this understanding appear from the second half of the second year on. Perhaps the best example is the development of empathy (e.g., Hoffman 1977). Younger infants, on perceiving another's exhibition of distress, will typically get distressed themselves and seek comfort for themselves. As the second year proceeds, infants in the same situation, while showing some distress themselves, will try to do something to bring comfort to the other. It is at this point, therefore, that they appear to understand that the feeling belongs to the other. In understanding the other, unlike in understanding the self, it is the first person information that has to be imagined and integrated with the immediately available third person information of the other. We assume that in the most primitive or novel circumstances, in which the intentional relation of another is to be understood, the first person information of the other is imagined using the child's first person information of self as a model (Harris 1989; 1991). In other words, the imagined first person information of the other is essentially the same as the subject's own first person information, but marked as belonging to the other. For example, when viewing another's distress early in this period, the child empathically feels distress but attributes it to the other. With experience, however, it is likely that the understanding of intentional relations of others will no longer require an appropriate empathic response from the child. Rather, the child will recognize the third person information of the other, and the first person component will be imagined. In this way, young preschoolers can recognize the diversity of intentional orientation of self and other to a particular object in some circumstances. For example, Flavell et al. (1990) have shown that 3-year-olds, presented with a favorite cookie and then shown another person taking a bite of it and expressing disgust, can understand that even though they like the cookie, the other does not.

We are proposing here that the understanding of third person information of the self and the understanding of first person information of others both depend on the capacity for imagination. This proposal suggests that the understanding of self and other develop in parallel (an idea that goes back at least as far as Baldwin, 1894) and there is some empirical support for this (see also Perner 1991). A number of studies have shown a relationship between self-recognition and empathy (e.g., Bischof-Kohler 1988; Johnson 1982). Most recently, for example, Bischof-Kohler (1988) found a correlation, independent of age, between self-recognition behavior using the marked face paradigm and empathic behavior directed at a distressed adult in an experimental setting.

Although the imaginative abilities of stage 2 allow differentiation of the intentional relations of self and other, intentional understanding is by no means complete. The limitation of this stage is that, because of restrictions on the child's imaginative power (e.g., Olson 1993), the understanding of the intentional relations of both self and others involves combining one current perceptual component and one imagined component. Thus, while children at this stage can recognize diversity of intentional relations, their understanding is still tied to current experience.

4.3. Stage 3 (level 4): Children's theories of mind and false belief. Much recent research on social understanding during the preschool period has been conducted under the heading "the child's theory of mind" (see Astington et al. 1988; Frye & Moore 1991; Wellman 1990; Whitman 1991a). Rather than giving a comprehensive review of the research on children's theories of mind, we shall concentrate here on one central issue in this literature, namely the understanding of false belief. False belief has been considered a kind of litmus test (e.g., Chandler 1988; Wellman 1988) for the presence of a theory of mind since Premack and Woodruff's (1978) seminal work with chimpanzees. Premack and Woodruff's target article generated intense interest among commentators with respect to what would be required as firm evidence for a theory of mind (e.g., Bennett 1978; Dennett 1978). That interest spread into developmental psychology with the work of Wimmer and Perner (1983) who, responding to the suggestions of the commentators, developed the first false belief task.

In their task, Wimmer and Perner presented children with a story in which a character, "Max," places some chocolate in one cupboard in the kitchen and leaves. While
Maxi is away, another character removes the chocolate from the first cupboard, places it in a second cupboard, and then also leaves. Maxi then returns, and the child subject is asked to predict where he will look for the chocolate. The correct answer, of course, is to predict that he will look in the empty cupboard and not in the cupboard where the chocolate really is. Wimmer & Perner found that it was not until about 4 to 5 years of age that children performed this task correctly. Younger children typically chose the real location of the object. On the basis of this result and others showing a developmental shift in performance on various measures of false belief understanding at about 4 years of age (e.g., Gopnik & Astington 1988; Moore et al. 1990; Perner et al. 1987), some have argued that what develops at this age, as indexed by success on false belief tasks, is the understanding that beliefs are attitudes toward representations of reality rather than toward reality itself, or what has been termed “a representational theory of mind” (Aston & Gopnik 1991; Perner 1988; 1991).

While the finding that it is not until about age 4 that children understand false belief is a fairly robust one (but see Freeman et al. 1991; Mitchell & Lacohee 1991; Siegal & Beattie 1991), younger children do have some knowledge of belief–desire reasoning. Most importantly, perhaps, they appear to know a considerable amount about desire and true belief or knowledge. Wellman (1990) has provided the most extensive coverage of 3-year-old children’s belief–desire reasoning. Using brief story tasks, in which the subject is told a story about a character who wants something or thinks something, Wellman and his colleagues (Wellman & Bartsch 1988; Wellman & Woolley 1990) have shown, among other things, that 3-year-olds know that people will act on the basis of their desires and knowledge. As previously noted, 3-year-olds can also recognize the distinction of intentional relations between self and other to a particular object in some circumstances (Flavell et al. 1990).

The questions of theoretical importance to us here are: What makes the false belief task particularly difficult, and what accounts for the apparent change in social understanding at about 4 years of age? Let us consider what it means to fail the false belief task. This task requires children to predict what the story character will think or do; children younger than about 4, as noted, typically fail the standard version because they answer on the basis of the real location of the hidden object. But what is meant by “the real location” in this context? “The real location” clearly does not refer to the state of reality independent of any agent or observer, because one could imagine that if, unbeknownst to the 3-year-old subject, the object in the false belief task were moved again, that subject would still predict that the agent will search in the location where the subject believed the object to be. For the purposes of the task, “reality” essentially means “the content of the child subject’s own current belief” or the situation as it appears to the child through current first person information. There are, then, two critical aspects to note about the task. First, the task requires the child to generate a representation of the agent’s epistemic relation to the situation, for which that agent’s third person information is not perceptually given. Second, the task also requires the child to imagine the first person information of the agent’s intentional relation, while the first person information about the child’s own intentional relation is different. Both these aspects are critical because together they ensure that neither current third person information nor current first person information are available to generate the requisite representation of the intentional relation. As a result, both components of the agent’s to-be-predicted intentional relation must be imagined.

Our argument, then, is that the Wimmer and Perner (1983) false belief task and its variants (e.g., Perner et al. 1987) are particularly difficult because they require the subject to generate a representation of the intentional relation of an agent when neither the first nor third person informational component is provided by current experience. It is under this particular set of conditions that the younger child has difficulty because, as we have already suggested, younger children are able to imagine only one component of an intentional relation. When presented with perceptual third person information about others, they are able to imagine the coordinate first person information. However, in the absence of perceptual third person information about the other, the representation of the other’s intentional relation can only be generated by using the current first person information of self, and in doing so the children make the wrong prediction.

The tasks on which younger children show reasonably good abilities tend not to have both of the critical aspects just outlined. On the one hand, Flavell et al.’s (1990) task, which shows understanding of emotional diversity; presents the child with perceptual third person information about the other; hence the child need not generate a representation of the other’s intentional relation, which involves both imagined third person information and imagined first person information. The child can deal with the diversity of emotional orientation by comparing two attitudes, both of which involve one perceived and one imagined component. On the other hand, tasks assessing true belief or knowledge, while they may require imagining the third person information about the other, are relatively straightforward because the subjects’ use of their own current first person information to generate the representation of the intentional relation of the other will lead to an accurate prediction. Similarly, if one examines the research on desire, it can be argued that there have been few, if any, tasks used to assess desire understanding that are clearly analogous to the false belief task. In other words, in virtually no case have subjects been presented with a situation in which they must predict the desire of another person while they themselves hold a conflicting current desire. The only case that comes close to being an assessment of desire understanding in the presence of a current conflicting desire is a study by Gopnik and Slaughter (1991), in which children had to recall their own previous desire while holding a different current desire. In this task, there was no available perceptual information, of either first or third person kind, on which to base an answer to the critical test question. Half of the 3-year-olds failed this task, answering that they had previously held the desire they now held.

If our argument so far is correct, then what develops at about 4 years is the capacity to perform what amounts to a double imaginative act, whereby the child can generate, through the use of the intentional schema, a representation of an intentional relation for which both first and third person information are imagined. As with the previous stage, we believe that the change in social understanding at about 4 years is made possible by a domain-general cogni-
5. Pathology of Social Understanding: Autism

In recent years, there has grown a large literature assessing the possibility that autism is a kind of tragic "null phenotype" with respect to social understanding. Autism is a severe childhood psychopathology with onset prior to 3 years. It is typically characterized as involving three main behavioral abnormalities: impaired social behavior; symbolic disability, including language impairment and lack of symbolic play; and a strong need for stereotyped routines in daily life. On the basis of recent research from the theory-of-mind perspective demonstrating that autistic children perform considerably worse than mentally retarded controls on a variety of tasks assessing an understanding of other minds, a number of authors have proposed that autism is a relatively specific deficit with respect to the understanding of other minds (Baron-Cohen 1991; Baron-Cohen et al. 1985; 1986; Frith 1989; Harris 1989; 1991; Hobson 1989; 1990; 1993; Leslie 1987; 1988; Leslie & Frith 1990).

It is clear, however, that the problems of autism do not commence at 3 to 4 years, when normal children have a concept of mental agent and start to pass false belief tasks. Retrospective studies suggest that, on average, autism becomes noticeable to parents as an abnormality in social behavior during the first 2 years of life (e.g., Short & Schopler 1988). Autistic children do not show social pretend play, a type of behavior that emerges in normal children during the second year of life (e.g., Sigman & Ungerer 1984; Wing et al. 1977). Other research has shown that autistic children show abnormalities in behaviors such as joint attention, which may well be precursors to the difficulties seen in standard theory of mind tasks (Loveland & Landry 1986; Sigman et al. 1986). In a similar vein, autistic children produce protoimperative but not proto-declarative communicative acts (Baron-Cohen 1989; Curcio 1978; Landry & Loveland 1989; Mundy et al. 1990), showing that, although they express their own desires, they do not act as though they recognize that there is a potential for joint attention or that others have intentions.

Imitation has also been shown to be impaired in autism and this impairment may point to an even earlier origin of the autistic problem (for a review, see Smith & Bryson 1994). Dawson and Adams (1984) found that about half of their autistic subjects were operating imitatively at stages 2 and 3 of the sensorimotor period, stages that are reached in normal children at about 2 to 4 months (Piaget 1962). When attempts to imitate are seen in autistic children, they may show some curious abnormalities. For example, Ohta (1987) found a pattern of "partial imitation" of manual gestures in a substantial proportion of autistic subjects. In these cases, the subjects, who were seated face-to-face with the model, produced a gesture that reversed the orientation of the hands so that the visual appearance of their own hands most closely resembled the model's hands. Such attempts at imitation, therefore, appeared to be controlled only by visual information.

Other authors have supposed that autism is primarily a deficit existing in the affective domain (Kanner 1943; Hobson 1989; 1990; 1993; Mundy & Sigman 1989). For example, Hobson (1993) marshals a number of kinds of evidence in support of autism having origins in affective interpersonal relatedness. First, clinicians since Kanner's (1943) original work have described how interactions with autistic children seem to lack affective contact. Such clinical impressions are supported by empirical work that shows that autistic children's emotional expressiveness, both vocal and facial, in the company of others is abnormal (Dawson et al. 1990; Kasari et al. 1990; Yirmiya et al. 1989). In addition, autistic children are significantly worse than control children in categorizing and determining the meaning of emotional expressions (Hobson 1986; Hobson et al. 1989).

On the basis of these varied results, we suggest that if there is a single simple deficit responsible for the difficulties in social understanding seen in autism, it is probably that first and third person information are not integrated in the representation of own and other agents' intentional relations. This inability exists at the most basic level; in other words, autistic individuals never attain an understanding of intentional relations at level 2 of our model. The failure of social understanding to proceed beyond level 1 of our framework could occur for a number of reasons. As we argued in previous sections, the representation of intentional relations between agents and objects at level 2 requires that the individual be able to integrate matched first and third person information with respect to objects. Given that the occurrences of matched states in interactions in early infancy are largely maintained by the adult rather than the infant (Schaffer 1984), it seems likely that matching of first and third person information does occur in the early developmental contexts of autistic children. Thus, it is more likely that the problem is with the extent to which the autistic infant attends to and processes such matched information. In other words, the intermodal integration of first and third person information is either disordered or absent. The extent of this information-integration problem is so far unknown. It may be specific to the processing of certain types of matched intentional information, for example, the epistemic information available in joint attention (Baron-Cohen 1981) or emotional information (Hobson 1989; 1990). However, given that the literature has implicated action-oriented, emotional, and epistemic deficits in autism, we suspect that the problem in autism is a general one with respect to the integrated representation of matched first and third person information. Further, given that autistic children appear disordered in very basic affective and imitative abilities (Dawson & Adams 1994; Hobson 1993; Ohta 1987), we suggest that the problem is not
limited to the integration of first and third person information about intentional relations, but affects all intermodal social information processing, even that prior to the stage when objects are normally brought into the interaction (see also Rogers & Pennington 1991). It may even be that the problem is a very general one in information processing, and is not limited to social information at all. Given that there is some evidence about the problems that autistic individuals suffer in the integration of information from more than one modality, such as vision and audition (e.g., Bryson 1970; 1972), it is entirely possible that autism is the result of a pervasive problem in integrating information across any modalities. Such a general difficulty with intermodal integration may itself be the product of an attentional impairment that prevents the relatively rapid shifts of attention across sensory modalities that are required for bimodal integration of information (Courchesne et al. 1993).

Whether the problem with information integration is a general one or is limited to social and perhaps just intentional information, it is relatively straightforward to see, from our account, how such an information-processing deficit would generate the severe problems in social understanding that occur in autism. As a result of this failure to integrate first and third person information, the autistic person is unable to enter into the shared intentional experience of level 2, normally evident in infants toward the end of their first year in behavior such as joint attention, and protodeclarative communicative acts. This would not interfere with the representation of their own desires from a first person perspective as well as the actions of others from a third person perspective. It would not, therefore, affect their acquisition of protodeclarative communicative skills, although the latter would tend to be quite primitive, taking advantage only of the mechanical causal properties of the other, for example, pulling or dragging a person. As an inevitable result of failure to integrate first and third person information, autistic individuals are unable to imagine intentional relations as properties of separate agents, including both self and other, at level 3 of our framework. Furthermore, although they may be able to recognize differences in knowledge among different others (Kazak 1992) on the basis of third person information alone, the representational theory of mind characteristic of level 4, which requires imagining divergent epistemic relations across self and other, would remain beyond the ability of the person with autism. Despite the sophistication of the knowledge that autistic individuals acquire about people, including themselves, the form of their representations of people remains entirely of either a first or third person sort.

Although according to our framework the autistic person uses only level 1 representations, we are not suggesting that their representation of the intentional relations of self or other is limited to the same range as that found in lower organisms. We assume, for example, that their ability to use the imagination is intact or at least in line with their general level of mental retardation. The use of the imagination would support abilities such as mirror self-recognition (see Mitchell 1993), which develops in autistic individuals at approximately the same time as in mentally retarded controls (Neuman & Hill 1978; Spiker & Ricks 1984). Furthermore, they acquire some use of language, which they apply to self and other (Frith 1989). Despite these indications of a lack of asymmetry in the understanding of intentional relations between self and other, however, we believe that such an asymmetry maintains itself. The autistic person acquires an ability to imagine first person information and generalize it to others, and to imagine third person information and generalize it to self, without ever integrating these informational components into a representation of intentional relations that can distinguish diversity in perspectives. A clear indication of this failure to integrate first and third person information is seen in their behavior in front of a mirror. Autistic individuals can imagine third person information about the self in order to recognize themselves in the mirror. However, they do not show any signs of self-conscious emotions (Neuman & Hill 1978; Spiker & Ricks 1984), implying a failure to integrate first and third person information for both the self and others. Consistent with this interpretation is the fact that autistic individuals refer to themselves as well as to others by name, but they rarely use personal pronouns, which, being deictic, require an appreciation of the intentional perspective of the speaker (Lee Hobson & Chiat 1994).

In sum, we believe that social understanding deficits in autism are the result of an inability to form representations of intentional relations with integrated first and third person components. This inability, we believe, is a general one, so that the understanding of both own and others’ epistemic, affective, and action intentional relations is disordered. The origin of the disorder appears to be impairment of the autistic individual’s ability to attend to or to integrate information across sensory modalities so as to extract information about perceptual invariants and build representations using such invariants.

6. Conclusion

In this target article, we have presented a theory of social understanding and the forms of representation of intentional relations. The theory describes a system of such forms that can be applied uniformly to self and other; it requires the use of an intentional schema that can integrate first person information about an intentional relation with third person information about that relation. We have suggested a four-level framework that describes the various forms of representation that do and do not require the intentional schema. The framework was applied to empirical evidence on the phylogeny and ontogeny of social understanding, and on human autism.

Although some aspects of our theory will no doubt be disconfirmed, we believe that it addresses certain issues in social understanding that are not, in themselves, open to disconfirmation. For example, it is a necessity of organismic existence that one has a qualitatively different kind of information about one’s own intentional relations and the intentional relations of others. Nevertheless, human beings have developed forms of representation of intentional relations that can be applied almost uniformly to self and other. How is this possible?

The current literature offers three apparent answers to this question, none of which seems to us entirely satisfactory. Those who postulate a “theory of mind” (e.g., Gopnik 1993; Perner 1991) succeed in accounting for the uniformity in applying the theory to self and other, but at the expense of providing a coherent account of how the theory takes hold of the first and third person information about intentional relations (Moore & Barresi 1993). Those who
postulate a process of imaginative simulation to account for social understanding typically focus on using one's own mind in simulating the mind of the other (e.g., Goldman 1989; 1992; Harris 1989; 1991; Humphrey 1984), but they do not make it clear how one can know one's own mind as such. Other stimulation theorists suggest that we acquire knowledge about both self and other through simulation (e.g., Gordon 1986; 1987; 1992), but they are not explicit on how this understanding comes about, especially in its original nonimaginative form. A third view, which postulates an original empathic relation between human persons (e.g., Hobson 1990; 1991), although dealing with origins of social understanding, fails to be explicit about its central mechanism of empathic relatedness.

Our proposal contains aspects of all three of these approaches without reducing to any one of them. We suggest that in human beings there has evolved a capacity to enter into shared intentional relations with another and an associated intentional schema to integrate first and third person information about these intentional relations. The resulting representations could be applied uniformly to self and other in comparable situations where only part of the informational source is available perceptually. This is one possible solution to the problem; there may be others. In any event, once knowledge of particular intentional relations has been acquired, the door is open to use the imagination and folk theories to further expand the understanding of intentional relations where only one, and eventually where no, direct input is available. At root, however, such understanding will always involve aspects of experience that have their source in the first person information about our selves and the third person information about others, and our ability to integrate these two forms of information into a single representation. This capacity to integrate knowledge about self and other is the basis for human social understanding and our appreciation for all aspects of the human condition.

ACKNOWLEDGMENTS
Preparation of this paper was supported by a grant from the Natural Sciences and Engineering Council of Canada to the first author and Grant No. 410-92-0694 from the Social Sciences and Humanities Research Council of Canada to the second author. Our thanks to Simon Baron-Cohen, Mark Bickhard, Robert Gordon, Paul Harris, Peter Hobson, Ray Martin, Terry Tomkow, Tom Vinci, Henry Wellman, and many colleagues in the Psychology department at Dalhousie for comments on earlier versions of the manuscript. We also wish to acknowledge the many useful suggestions for revision provided by the reviewers and editor of this journal.

NOTES
1. The term "intentional relation" derives in part from the concept of intentionality. Intentionality is a property of mental activities that has been described in various ways since Brentano (1874/1973) brought it into modern usage in an attempt to distinguish mental phenomena from physical phenomena (see Barre 1989; Chisholm 1967; Dreyfus & Hall 1984). Brentano's definition of intentionality focused on the intentional "inexistence" of the objects of mental acts, by which he meant that mental acts are "directed toward," "make reference to," or are "about" objects that need not exist in reality but exist only as objects contained within these acts. These "inexistent" objects occur in mental phenomena in various modes, such as being thought about or being desired. In the present paper we focus not on intentionality per se, but on intentional relations. In our terminology, intentional relations exist between an organism and actual or possible objects or states of affairs. They depend on sensorimotor and higher mental capacities of the organism and have their roots in certain activities that can be defined in nonintentional causal terms (see, e.g., Agar 1993; Dretske 1981; 1988; Fodor 1987; 1990; Millikan 1989; Shapiro 1992). Our presupposition is that any empirical account of social understanding must begin with how organisms appreciate intentional relations involving agents' activities directed at real rather than at inextant objects and that the latter understanding must depend on the former. Hence we begin with intentional relations rather than with intentionality.

2. This argument is rooted in the philosophical tradition of Merleau-Ponty (1964), Nagel (1974), Scheler (1954), Searle (1992), Strawson (1959), and Wittgenstein (1958). Like ourselves, these authors were concerned with the problem of how our knowledge of mental activities of self and others can be based on two qualitatively different kinds of information yet apply equally to self and other. Their answers differ from each other's and ours, yet all presuppose that there are two distinct kinds of information about mental activities. For instance, Nagel (1974) seems to have the view that there are two distinct forms of representation of mental activities, subjective and objective, but that both of these can be applied to self and other. In contrast, Wittgenstein (1958) seems to believe that there is a single linguistic form of representation for mental activities, but that our criteria for the application of this form is different when we apply it to ourselves and others.

3. Although we are claiming that there is uniform conceptual system for understanding the intentional relations of self and other, despite the differences in informational sources, this is not to say that there are no differences between understanding self and understanding others. Research in social cognition has revealed a number of instances of such differences, and some of these instances provide clear support for the idea that first and third person information tend to focus on different aspects of the intentional relation. One such example is the actor–observer effect (Jones & Nisbett 1972). In this effect, attributions of the causes of one's own behavior tend to focus on the situation that one is in, in other words, on the objects of one's intentional relations, whereas attributions of the causes of others' behavior tends to focus on the characteristics of the person acting. Thus, attributions for self tend to inherit the quality of first person information and those for others tend to inherit the quality of third person information.