2009

Understanding Norms Without a Theory of Mind

Kristin Andrews
York University

Recommended Citation

Andrews, K. (2009). Understanding norms without a theory of mind. Inquiry, 52(5), 433-448.

Understanding Norms Without a Theory of Mind

KRISTIN ANDREWS
York University, Canada

ABSTRACT

I argue that having a theory of mind requires having at least implicit knowledge of the norms of the community, and that an implicit understanding of the normative is what drives the development of a theory of mind. This conclusion is defended by two arguments. First I argue that a theory of mind likely did not develop in order to predict behavior, because before individuals can use propositional attitudes to predict behavior, they have to be able to use them in explanations of behavior. Rather, I suggest that the need to explain behavior in terms of reasons is the primary function of a theory of mind. I further argue that in order to be motivated to offer explanations of behavior, one must have at least an implicit understanding of appropriate behavior, which implies at least an implicit understanding of norms. The second argument looks at three cases of nonhuman animal societies that appear to operate within a system of norms. While there is no evidence that any species other than humans have a theory of mind, there is evidence that other species have sensitivity to the normative. Finally, I propose an explanation for the priority of norms over a theory of mind: given an understanding of norms in a society, and the ability to recognize and sanction violations, there developed a need to understand actions that violated the norms, and such explanations could only be given in terms of a person's reasons. There is a significant benefit to being able to explain behavior that violates norms, because explanations of the right sort can also serve to justify behavior.

There has been a growing interest in the idea that folk psychology and moral psychology are closely connected. A first answer to the question about the relationship between folk psychology and the moral sense might be that in order to understand moral norms, one must have a theory of mind. To have a theory of mind is to have the ability to attribute beliefs and desires to others, to use these mental state concepts for making predictions. Insofar as moral theories require that knowing the right thing to do requires knowing how an action will affect others, it is perhaps *prima facie* plausible that moral psychology relies on a theory of mind.

While I agree that there is a relationship between the moral sense and folk psychology, I disagree with the account sketched above. Recent research from the social and biological sciences can be seen as suggesting a different kind of relationship between these two abilities. Given findings from both human and animal studies, I believe that an implicit understanding of the normative drives the development of a theory of mind, and that a full-fledged theory of mind, which includes the ability to attribute beliefs, is not necessary for understanding at least some norms. Since understanding norms in some way is an essential element to having moral sense, this conclusion allows that moral norms may be among those that one can grasp without a theory of mind. Indeed, if one accepts the view that all norms are to some extent moral norms, then the conclusion becomes even stronger. In this paper I hope to offer some

reasons for thinking that moral reasoning drives the development of folk psychology, rather than the other way around.

The argument begins with the premise that a theory of mind (understood narrowly as the ability to attribute to an agent her primary reason for acting, consisting of both her belief and pro-attitude) is not necessary to predict the behavior of individuals who lack a theory of mind, and is not even the predominant method used to predict adult human behavior. In the theory of mind literature, it is often assumed that we need to know an actor's primary reason in order to predict behavior. The debates between simulation theory, theory theory, and hybrids, as well as the debates about theory of mind in other species, all seem to take for granted the notion that we predict behavior by attributing a belief and desire to the actor. I think there are some good reasons to reject this assumption. For one, there are a number of other very successful methods of predicting behavior. Also, young children appear able to make predictions before they develop a full-blown theory of mind. Finally, many nonhuman species also appear to predict the behavior of their conspecifics, but do not demonstrate the ability to attribute both beliefs and desires.

I have argued elsewhere that humans do not require facility with a theory of mind in order to predict behavior in normal situations (Andrews, 2008). I will begin by outlining a few reasons for that conclusion, and from there argue that the ability to explain behavior is necessary in order to have a theory of mind ability. I think that the primary role for theory of mind is explanation, not prediction, and I aim to show how this view can flip the traditional relationship between folk psychology and moral psychology.

I. Predicting without a theory of mind in normal situations

Predicting behavior is a quotidian activity for humans and members of other species. In most instances, when we predict behavior we don't even notice that we're making a prediction. The expectation that the others around us will act in certain ways is an important part of any functional society, and it allows us to go about our day without wondering if our co-workers will do their job or if commuters will drive within certain parameters. In these and other normal circumstances, we need not appeal to the actor's primary reason for her behavior. Instead, there are other methods we might use, some of which are also available to infants and nonhuman animals.

Let me briefly outline four methods adult humans use to predict behavior in such situations. We predict from self, expecting that other people will act as we do. For example, if we are forced to decide whether another person likes poetry, and we know little about that person, we are likely to attribute to her the same preference we have ourselves, automatically or without reflection (Krueger, 1998; Krueger and Clement, 1994; Marks and Miller, 1987; Mullen *et al*., 1985). This view is distinct from the usual understanding of what simulation theory requires for predicting behavior, because when we predict from self we rely on an epistemic egoism that makes it difficult for us to set aside privileged information (Royzman, Cassidy, and Baron, 2003). Despite this problem, there is reason to think that predicting from self is a useful strategy. Dawes and Mulford have argued that when we know nothing relevant about a person, a rational response is to use oneself as the sample, especially if the target shares some property with the predictor, such as being a fellow countryman or student (Dawes and Mulford, 1996). This is because one is more likely to make an accurate prediction by using one's self as a model in such circumstances.

We predict from the attribution of stereotypes or social norms, expecting that the woman will clean the kitchen or that the waiter will take our order. People who fit into a clear category, either in the case of a race/gender/class/subculture stereotype, or in the case of an established social role, are expected to behave in ways that we have come to associate with that group (Locksley *et al*., 1980; Locksley *et al*.,

1982). Children begin using gender stereotypes at a very early age. Some researchers suggest that at 2 years children are already shaping their gender identity and developing an own-sex schema that they use to categorize activities as appropriate or not (Martin, 2000).

When we have more knowledge about an individual's past behavior, we predict from inductive generalizations. Young children are very good at making inductive generalizations over biological or physical properties of humans and other objects (Carey, 1985; Gelman, 1988) and at an early age they are able to accurately predict behaviors based on idiosyncratic physical attributes such as hearing acuity (Kalish, 2002). By age 10, children are making inferences about the future based on past observations, and will judge that a character who performed a kind action on one occasion is likely to behave kindly in a new situation (Rholes and Ruble, 1984).

We also make predictions based on a person's trait or temperament (Ross and Nisbett, 1991). We may learn of a person's trait either through the observation of that person's past behavior, or by having the person's trait described to us. We use trait attributions to make predictions when we don't have knowledge of a person's past behavior in the specific situation. Our inference from observing past behavior to a character trait (for example, from observing a colleague offering to chair a committee to concluding that the colleague is helpful) leads us to make predictions about that person's helpful behavior in other circumstances as well (e.g., to predict that the person will offer to take on extra household chores).

Young children who do not yet have the ability to ascribe to others their primary reasons are also able to predict behavior; even in infancy, children are able to make many predictions. Long before humans develop the cognitive abilities necessary for ascribing beliefs or even personality traits, they are able to predict behavior based on what Trevarthen calls "primary intersubjectivity" (Trevarthen, 1979). In the first two years of life, long before they first pass false belief tasks, children make predictions when they imitate, engage in joint attention and triadic interactions, and respond emotionally to facial expressions and speech prosody. Young children are able to make predictions by attributing mental states without knowing an actor's primary reason; children are sensitive to others' emotions, intentions, and perceptual states from an early age. For example, by 12 months an infant is able to predict that an actor who looks at an object with a positive facial expression is more likely to grasp that object than some other one (Phillips, Wellman, and Spelke, 2002). There is no evidence that children have an understanding of desire before 18 months, and it is thought that children do not develop an understanding of belief until about 4 years (Wellman *et al*., 2001), though it may be that children don't have a robust sense of belief that includes referential opacity until years later (Apperly and Robinson, 1998; 2003; Russell, 1987).[1]

Nonhuman animals too are able to predict behavior, though there is no evidence that members of any other species are able to predict behavior by ascribing primary reasons to others. Chimpanzees, for one, have been shown to make accurate predictions of human behavior. For example, a recent study by Felix Warneken and Michael Tomasello (2006) established that chimpanzees engage in helping behaviors in response to a human caregiver's nonverbal request for an out-of-reach object, as do 18-month-old children. The children, however, demonstrate a greater degree of helping behavior across task types than do the chimpanzees.

Chimpanzees, like human children, are also quite sensitive to emotional responses indicated by facial expressions. For example, Lisa Parr's research demonstrates that chimpanzees are able to categorize facial expressions associated with different emotional responses. Using a match-to-sample paradigm, Parr and colleagues have shown that chimpanzees recognize a number of different facial expressions (Parr, 2003). Following Ekman's work on emotion in human facial expressions, they have created a Chimpanzee FACS (Facial Action Coding System) that they use to construct models of chimpanzee

expressions in order to determine which configurations of muscle movements the chimpanzees find salient in their perception of emotion.

In all the cases described above, for both the humans and nonhumans, the kind of behavior that is being predicted is of a rather banal sort. To predict that someone will grab an object she is looking at, or that students will show up for class, are examples of predictions in normal contexts. While these points support the conclusion that a theory of mind is not needed to make predictions in such quotidian cases, it doesn't follow that we never need a theory of mind to make predictions of behavior. Indeed, some situations require it.

II. Predicting behavior in an anomalous situation

While it is relatively easy to predict what someone will do in ordinary contexts, the case is quite different when the situation is novel in some way. For example, when a quotidian prediction turns out wrong, we are then in an anomalous situation, and predicting what others will do next requires something more than relying on regularities previously observed—appealing to our past experience or the actor's past behavior will not help, nor will recognition of the individual's traits or stereotypic properties. A truly anomalous situation does not permit predictions using any of the methods discussed above.

For example, you see your normally calm and sensible colleague yelling obscenities at a girl on the street, and you wonder what he would do if you were to approach him. None of the methods discussed previously will be of much use here, because your colleague is acting so out of character. In order to figure out whether your colleague would welcome your intervention and take you as an ally or turn and start yelling at you as well, determining his primary reasons for the action seems to be the only recourse. When we see an in-group member engaged in behavior that appears incomprehensible, our initial response is to try to make sense of it. We reconstruct the situation in such a way as to make the incomprehensible comprehensible. In order to do so, we construct a story having to do with the actor's reasons for acting. To predict the colleague's response to your intervention, you have to decide whether your friend thinks that the girl is guilty of something, or whether he has a more general target for his anger and the girl is simply the unlucky recipient of his mood. That means you need to explain what he is doing right then, in order to predict what he is going to do next. Before you can predict, you must explain.

Not only do we need to offer explanations before we predict in anomalous situations, we need to offer explanations in terms of the actor's reasons. While reason explanations are perhaps the most discussed, we also explain people's behavior in other terms, such as in terms of the individual's past history or current situation (Malle, 2004), or even in terms of personality traits, physical or mental limitation, and so forth. However, the anomalous situations I am talking about are not only unpredictable in such terms, but are also unexplainable by them; the behavior is anomalous because these other kinds of explanations do not work. Anomalous situations are only explainable in terms of a person's reasons.

While observers more often explain behavior by appealing to the causal history of the individual, and actors more often explain their own behavior by appeal to beliefs and desires, there are some conditions in which observers do explain others' behavior via belief/desire attribution (Malle, 2004; Malle *et al*., 2007). Malle has found that observers will attribute beliefs and desires to actors when they are motivated to portray the behavior in a positive light. I suggest that when a group member acts in an anomalous manner, others in that group will be motivated to portray that behavior in a positive light, and hence the explanations that are generated will tend to be reason explanations in terms of propositional attitudes.

If this reasoning is sound, then it follows that to predict behavior in an anomalous situation, one requires a theory of mind. But it also follows that to predict behavior in such a situation, one must *first* be able to

explain behavior in terms of an actor's primary reasons. That is, theory of mind prediction rests on a prior ability to engage in theory of mind explanation.

One might object here that while a theory of mind is not needed to make the kind of quotidian predictions discussed in the previous section, it may still be the case that humans do use a theory of mind to make these predictions. I am suggesting that theory of mind developed in order to explain behavior, but perhaps theory of mind developed independently from our predictive and explanatory practices. If, having already developed the ability to attribute beliefs and pro-attitudes, people are able to then use a theory of mind to predict normal behavior before they are able to make predictions in anomalous situations, then theory of mind prediction may in fact be developmentally prior to theory of mind explanation.

The primacy of theory of mind explanation is key to my argument here, and I think that the objection is not terribly serious, given the lack of evidence in favor of it. The burden of proof rests with the objector, since here is no independent pressure toward developing a theory of mind for making quotidian predictions, and creatures who lack a theory of mind (including young children) are quite successful at making predictions nonetheless.

Another point in favor of my premise that theory of mind is likely not used to make quotidian predictions has to do with the relative accuracy of theory of mind predictions and predictions made using the other methods. First, theory of mind predictions suffer from the underdetermination of reasons for any one behavior. Perhaps young John goes to church with his mother to make her happy, or perhaps he goes because he hopes to avoid a fiery afterlife; both reasons are consistent with his action. It will therefore be difficult to correctly identify the reason a person has for acting without, perhaps, asking for those reasons. Further, by considering a person's beliefs and desires we may actually undermine our chances at accuracy because prediction via belief/desire attribution involves consideration of the reasons our target has for behaving in a certain way, and it has been found that by considering someone's reasons for action, we come to see the action as more likely (Wilson and LaFleur, 1995). For example, when college students were asked to predict whether they would act in a friendly or unfriendly way toward another student, Wilson and LaFleur found that participants who were asked to consider their reasons for acting subsequently judged their prediction as more likely than a control group. Wilson and LaFleur also measured the participants' actual behavior, and they found that the experimental group made less accurate predictions. This effect is explained in terms of confirmation bias. In general, when we examine whether a theory is correct we use a positive test strategy, and attend to the information that makes the hypothesis seem more likely. Thus, in developing a theory, we are likely to accept the first plausible explanation, and act on that.

When we attempt to predict behavior based on a belief/desire set that is consistent with the situation, we should be expected to make the same error. Because there are many different belief/desire sets consistent with any one situation, and belief/desire sets are paradigmatic reasons for others' behavior, these findings suggest that we would likely fixate on the *first* belief/desire set we generate, rather than the *best* belief/desire set available. These problems with accuracy when making theory of mind predictions are in contrast to the general accuracy we see in our quotidian predictions.

Taking these points into consideration, I will accept the premise that there is a close relationship between analyzing anomalous situations and the development of theory of mind ability. That is, I think that the drive to explain behavior is what leads us to develop a theory of mind. If I am right about this, then rationalizations, not predictions, are the original home for theory of mind ability.

III. Recognition of a situation as anomalous implies an understanding of norms

It is at this point that I can begin to present the argument that normative understanding is prior to theory of mind understanding. Given the premise that the ability to explain behavior develops before or with the development of theory of mind, I want to argue that at least an implicit understanding of normative rules is prior to a theory of mind.

Since the construal of behaviors as caused by beliefs and desires requires seeking an explanation of behavior, and explanations are sought in contexts where the behavior is construed as anomalous, bizarre, or inappropriate, agents must have some understanding of what counts as anomalous behavior. In order to see a piece of behavior as anomalous, one must have a background expectation about normal behavior. It is this background expectation that I take to demonstrate at least an implicit understanding of norms. By norms, I am referring to societal rules and expectations; at this point I need not defend the claim that these are, strictly speaking, moral norms, although some of them may be.

Of course, it is not the case that every time one witnesses an exception to a statistical regularity, one thereby witnesses the violation of a societal norm. This is true even if the exception to the statistical regularity in question involves agents; for example, there are statistical regularities associated with handedness in humans, but one's first observation of someone preferring her left hand need not drive the asking of a why question. If there is no motivation to ask a why question answerable in terms of reasons, then on my view the exception to the statistical regularity is not a violation of a norm. This is enough to show that statistical regularities and norms are different even if the norms in question are not, strictly speaking, moral norms. Only salient violations of regularities will lead one to seek understanding of the behavior, and the construal of a behavior as a salient exception to a statistical regularity suggests at least an implicit understanding of what is normal. The kinds of expectations whose violation motivates explanation-seeking behavior, where the explanation sought is in terms of a person's reasons, will be some subset of anomalous behaviors, namely those violations that have consequences in ethics or something else of import for individuals in the society. Construal of a situation as an exception to such regularities implies an understanding of appropriate behavior, I suggest, and that implies an understanding of norms.

Of course, in some cases when you see someone fail to engage in a typical behavior, your explanation need not appeal to the person's reasons. It may be obvious that the person is incapable of engaging in the action due to physical or mental limitations. For example, there is no mystery as to why the fatally ill person fails to play basketball as she usually does; one can easily explain that in terms of capability. Though the situation may be anomalous in the individual case (since Sue always plays basketball), it fits into a larger generalization about what sick people can and cannot do. It is normal for sick people to avoid physical activity, and Sue now falls under the sick person stereotype.

One who understands some behavioral patterns as normal must have at least a procedural knowledge regarding normal or appropriate behavior. This goes both for an individual—knowing Sue's normal behavior—and for people generally—knowing normal adult behavior. The rules that describe such behavior may not be declaratively represented; they may take the form of some implicit knowledge *how* rather than propositional knowledge *that*. One shouldn't object that such behavior doesn't indicate understanding of norms until there is meta-awareness of such norms, for such an objection would also lead to the conclusion that most humans lack moral knowledge, and this conclusion is *prima facie* false. We see that adult humans don't appear to have privileged access to all the norms they use when making moral judgments on trolley problems, for example (Hauser *et al*., 2007).

The argument in support of the view that one understands norms before developing a theory of mind can be sketched as follows:

1. The ability to explain behavior in terms of reasons develops with the development of theory of mind.
2. To develop the ability to explain behavior in terms of reasons, one first has to be able to construe a situation as the exception to a certain kind of regularity.
3. Construal of a situation as such an exception implies an understanding of appropriate behavior.
4. To understand behavior as appropriate is to understand norms.
5. Therefore, at least an implicit understanding of norms is prior to a theory of mind.

This first pass at defending the claim that some understanding of norms is prior to the development of theory of mind is based on a host of empirical studies on children and adults, findings that challenge the traditional relationship between understanding norms and understanding reasons. The argument is merely suggestive, given that I haven't provided a knock-down argument for the premises. However, I hope to have at least established their plausibility. To bolster the conclusion, I will provide another sort of argument by looking toward evidence in other species that lack theory of mind ability but exhibit normative understanding. This argument will be presented in the next section.

IV. Norms in animal societies

There is a large body of data on the norms of primate societies, especially within the research projects investigating cultural differences within primate species. The culture project is relevant to this discussion because culture includes not only the way we do things (McGrew, 1978; 1992) but also the standards of one's community (McGrew, 2009). Cultural traditions imply the existence of implicit codes of behavior, because they refer to population-specific differences in behavior that (a) are not ascribable to purely ecological differences between communities (b) are socially learned and (c) persevere for some time. Collaborations between researchers studying primates in the wild led to the discovery of a number of behaviors that appear to fit the criteria for cultural differences in chimpanzees (Whiten *et al*., 1999), orangutans (Van Schaik *et al*., 2003), and capuchin monkeys (Perry *et al*., 2003). For example, the chimpanzee project has designated 571 candidate behaviors as culturally variant (Whiten, 2007). Many of the behaviors on Whiten's list are related to different methods of food processing, which we wouldn't expect to reflect an interesting norm that is emotionally salient to these animals. Because I am ultimately concerned with norms that may be related to moral agency, I will focus here on the cultural behaviors that appear to have purely social benefits. Let me provide three examples in two different species of primate.

The primatologist Susan Perry suggests, following Zahavi (1977), that some of the social traditions she has observed among capuchin monkeys may have the function of communicating trust and strengthening bonds:

> All of these behaviors occur in relaxed social contests, in which the participants are typically somewhat isolated from the rest of the group . . . the participants move slowly (which is highly unusual for a capuchin) and have trancelike expressions on their faces, staring out into space or else looking at their fingers . . . all of them [the activities] involve a certain amount of risk or discomfort to one or both participants. Hand-sniffers have one another's fingernails delicately lodged in their nostrils, which restricts their movements; tail-sucking and the finger-in-mouth game involve placing a body part between the sharp teeth of the partner (since many capuchins are missing digits and tail tips, it is reasonable to assume that this is risky); the hair-passing game involves yanking significant amounts of hair out of the face and shoulders, which cannot be very comfortable. Perhaps these

> conventions are ways of testing the bonds between individuals . . . (Perry *et al*., 2003, p. 254)

Perry reports that the introduction of this behavior to a new monkey is a time-consuming process; the initiator slowly develops a relationship with another monkey, and the two monkeys negotiate their physical interaction, culminating with each monkey inserting its fingers deep into the nostril or even the eye socket of the other. There appear to be rules to these behaviors designed to avoid causing harm, and these rules were developed cooperatively by the initiator, a young subordinate male named Guapo, and the other monkeys with whom he engaged in this behavior.

Handsniffing and the other games may be an example of a developed social norm in these capuchin societies. Perry suggests that all these behaviors involve trust, that these rituals may be used to signal commitment to the social relationship, and that failure to perform the ritual may have long-term negative consequences for the relationship. Individuals must coordinate with one another to touch in ways that involve risk, and in ways that are made available only to an exclusive subset within the community. In addition, the games involve turn-taking and role-switching, suggesting that there are well-developed rules to follow. If the norm is violated, a capuchin will be injured, loosing a finger-tip or sustaining damage to an eye. Moreover, if the social bonding interpretation is correct, violations could also lead to emotional distress over the deterioration of the bond.

Another norm, this one seen in chimpanzee societies, has to do with territorial behavior. Chimpanzees are one of the few species known to form coalitions to engage in large inter-group hostile encounters. Such aggression has led to the extermination of one known chimpanzee community at Gombe (Goodall *et al*., 1979). When males form patrol groups, their behavior changes dramatically. John Mitani describes the behavior: “Males are silent, tense, and wary. They move in a tight file, often pause to look and listen, sometimes sniff the ground, and show great interest in chimpanzee nests, dung, and feeding remains” (Mitani, 2002, p. 18).

These patrols move along the periphery of their territory, sometimes making incursions into neighboring territories in order to hunt colobus monkeys, and often they run across members of the other community. Depending on the size of the patrolling group, and the size of the group they encounter, the patrol may either back away quietly, or attack.

Mitani reports that chimpanzees from Ngogo in Kibale National Park, Uganda will often hunt outside their territory, and when they do so they are silent and take prey back to their own territory very quickly. However, when inter-group encounters result, the Ngogo chimpanzees have been observed to kill and eat infants from the neighboring community. They have also been known to attack and kill adult males, whom they castrate rather than eat. Mitani suggests that the Ngogo chimpanzees might have started forming hunting patrols that take them outside their own territory because there are few monkeys left in their own territory; the Ngogo community is quite large, and they appear to have over-hunted.

While describing this behavior as warfare over scarce resources might be an over-attribution, the behavior reflects the norms and standards of the chimpanzee society, standards that reflect ideas of property and responses to violations of property norms. The Ngogo chimpanzees are very successful in the cross-boundary incursions, winning most of the battles. Despite this, they continue to go into patrol mode when they cross the border.

One additional example: some chimpanzee societies engage in a highly complex cooperative hunting strategy, and have meat-sharing rules corresponding to the individual roles performed by those in the hunting party (Boesch, 2002). Typically, there are four roles that the animals will take when hunting

monkeys: driver, chaser, ambusher, and captor. When the prey is spotted, each of the hunters takes on one of these roles, based on their location in relation to the monkey and their anticipation of the monkey's behavior. The hunters have to behave flexibly, for they will change roles as the situation dictates, and fall back to rely on one another if that seems to be the most efficient way to achieve the goal. Each of these roles is quite sophisticated, and it can take the chimpanzees twenty years to become proficient in the more sophisticated hunting roles.

Once the hunt is concluded, the meat is divided up between the four hunters. While the age and dominance of each member of the hunting party affects the distribution of meat, the most significant factor determining distribution is the degree of effort. The meat-sharing rules state that the largest share of the meat goes to the animal who did the most to catch the prey. As an implicit rule governing behavior, chimpanzee meat-sharing may be seen as a norm that deals with fair distribution and cooperation, and involves negotiating between one's personal desire for the meat and the more impersonal value associated with fair distribution.

These three examples of behavior in primate societies demonstrate sensitivity to societal norms and cultural expectation of rule-following. However, there is no evidence that these species are able to explain behavior in terms of an actor's primary reasons. While chimpanzees show sensitivity toward others' ability to see objects (Hare *et al*., 2001; 2006), and make predictions that may be seen as based on the attribution of some states, this ability falls short of full-fledged theory of mind.

While there isn't evidence that nonhuman primates have a theory of mind, they do have the ability to develop variations in their behavioral repertoire that involve creating, following, and violating social norms having to do with trust, harm, and cooperation. These primates appear to have societal norms, and individuals appear to have at least an implicit understanding of the relevant normative rules. Indeed, individuals across species demonstrate implicit understanding of norms as demonstrated by responses to violationsof those norms. Observational data suggests that individuals who violate social norms can suffer consequences such as social ostracization or attack (see Brosnan, 2006 for a review). That such animals also lack a theory of mind further supports the claim that normative understanding can exist without a theory of mind. Since they lack a theory of mind, they cannot explain violations of the norms in terms of reasons. The ability of these animals to follow norms and to punish violations without a theory of mind merely indicates that norms can exist prior to understanding other's beliefs and pro-attitudes. In the next section I will offer a hypothesis explaining why individuals develop the ability to recognize norms before they develop an understanding of reasons.

V. A proposed relationship between theory of mind and norms

I have offered two different arguments for thinking that individuals develop normative understanding before they develop a theory of mind. At the beginning of this paper I promised to draw a conclusion about the relationship between moral psychology and theory of mind. To do so, I need to suggest that some of the norms I have been discussing are moral norms. However, I want to avoid the quagmire that is metaethics, so instead of entering into the debate on the nature of norms, let me simply point out that some of these norms will be seen by some ethicists as moral ones. Norms of fairness, of harming and helping others, of justice, rights, cooperation, and so forth are often seen as moral norms. Indeed, it might be the case that all norms whose violations require an explanation in terms of reasons should count as moral norms, but I am not prepared to argue for that conclusion here.

If it is the case that individuals develop an understanding of norms before they develop a theory of mind, we might wonder whether there is a reason for this trajectory, which may exist phylogenetically as well as ontogenetically. Let me propose an answer. Given an understanding of norms in a society, and the ability

to recognize and sanction violations, there developed a need to understand actions that violated the norms. Explanations for norm-violating behavior that didn't cite a person's reasons either led to excluding the individual (e.g., "He *f*ed because he is crazy, so let's stop sharing meat with him"), or they failed to satisfy those who demand an explanation. This need to have a satisfactory reason for the behavior of one's companions is what drives the need to develop another sort of explanation, namely reason explanations. There is a significant benefit to being able to explain behavior that violates norms, because explanations of the right sort can also serve to justify behavior.

Of course, not all explanations are justifications, since we can provide the psychopath's reasons for his crimes without condoning his actions. But it is also the case that all justifications will refer to a person's reasons for action. A justification is a reason explanation plus an indication that the reasons are acceptable. Sometimes it sounds as though our justifications refer to the situation rather than to the actor's reasons (e.g., he left her because she was having an affair). However, if this is an actual explanation, the explanation must be known by the actor, and be among the actor's own reasons. Consider saying "Joe left Hilda because she was having an affair but he didn't know she was having an affair". This statement makes little sense. If Joe left Hilda because she was having an affair, then he knew she was has having the affair, and one might take his action as justified given the situation. However, if Joe left Hilda and she was having an affair, but Joe didn't know or didn't care that Hilda was having the affair, and instead left her because he was bored, then it is no longer correct to say that Joe left Hilda *because* of the affair. And correspondingly, one may no longer think that Joe's action was justifiable, given the reason that caused it.

Why is it beneficial to be able to justify behavior? For one, justifications of prohibitions help to refine norms, to point to exceptions to the norms or to help clear up grey areas. Consider a universal norm that one ought not steal. Heinz violates this norm by stealing medicine for his sick wife. Unless Heinz is able to give his reasons for the action, he will be subject to the society's sanction for the violation of the norm. However, if the society is able to understand and consider Heinz's reasons, then not only might it be beneficial to Heinz (insofar as he may avoid punishment) but it can also be beneficial to the society as a whole. The society benefits insofar as it can then create more nuanced norms, ones that better capture the motivation that led to the creation of the original norm in the first place.

But perhaps more importantly, justifying behavior can help a society develop altogether new norms that will provide greater benefits to individuals. For example, recognizing Heinz's reasons for stealing can lead a society to develop welfare systems, which in turn leads to the introduction of new norms. For a society to progress morally, new norms must be introduced. In order to introduce new norms, and to convince society to abide by those norms, one must give justifications for acting in that way. Justifications of behavior, as I argued above, will take the form of reason explanations. The ability to provide reason explanations will facilitate the introduction of new norms, and providing reason explanations requires a theory of mind.

If having a theory of mind requires first that one understands norms, then the cognitive requirements for (at least implicitly) understanding norms are rather lower than is commonly thought. Since moral norms are likely among the norms that can be understood without a theory of mind, moral knowledge would be possible without knowledge of beliefs and pro-attitudes. This conclusion makes tenable the claim that some animal societies are moral societies, and it should be seen as supporting research projects that look for moral knowledge among other species. However, this conclusion also points to a limitation in such societies. Moral growth, I suggest, will be hindered in societies without a theory of mind, since in order to change the moral norms in a rational way one must justify the new behaviors, and justifications will refer to an actor's reasons.

Living in a society with norms of behavior that are sometimes violated may drive one to seek an explanation for the violations. Once that question is asked, an individual is well on the way toward understanding that people act for reasons.

Note

1. While some claim infants as young as 15 months understand belief and have a theory of mind, I think that conclusion is not supported. In their study on infant theory of mind, Onishi and Baillargeon (2005) first familiarize their participants with a scene including an actor, an object, and two boxes. The actor then places an object in one of the boxes, and leaves the scene. While the actor is away, the object moves itself from one box to the other box. When the actor comes back, she looks into one of the boxes. Infants stared at the scene longer when the actor looked into the box that currently held the object, rather than the box where the actor left the object. While some interpret this finding as indicating that infants have some understanding of false belief, there are other possible interpretations of the results. For example, infants could be operating on behavioral rules such as "People look for things where they left them," or on some other low-level similarity to past observations (Ruffman and Perner, 2005).

References

Andrews, K. (2008) "It's in your nature: A pluralistic folk psychology", *Synthese*, 165, pp. 13–29.

Apperly, I. A., & Robinson, E. J. (1998) "Children's mental representation of referential relations", *Cognition*, 67, pp. 287–309.

Apperly, I. A., & Robinson, E. J. (2003) "When can children handle referential opacity? Evidence for systematic variation in 5- and 6-year-old children's reasoning about beliefs and belief reports", *Journal of Experimental Child Psychology*, 85, pp. 297–311.

Boesch, C. (2002) "Cooperative hunting roles among Taïe chimpanzees", *Human Nature*, 13,

pp. 27–46.

Brosnan, S. (2006) "Nonhuman species' reactions to inequity and their implications for fairness",

Social Justice Research, 19, pp. 153–85.

Carey, S. (1985) "Are children fundamentally different kinds of thinkers and learners than adults?" in: S. Chipman, J. Segal & R. Glaser (Eds.), *Thinking and Learning Skills*, 2, pp. 485–517 (Hillsdale, NJ: Erlbaum).

Dawes, R. M., & Mulford, M. (1996) "The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment?" *Organizational Behavior and Human Decision Processes*, 65, pp. 201–11.

Gelman, S. A. (1988) "The development of induction within natural kind and artifact Categories", *Cognitive Psychology*, 20, pp. 65–95.

Goodall, J., Bandora, A., Bergman, E., Busse, C., Matama, H., Mpongo, E., Pierce, A., & Riss, D. (1979) "Intercommunity interactions in the chimpanzee population of the Gombe National Park" in: D.

Hamburg & D. McCowen (Eds.), *The Great Apes*, pp. 13–54 (Menlo Park, CA: Benjamin/Cummings).

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000) "Chimpanzees know what conspecifics do and do not see." *Animal Behavior*, 59, pp. 771–85.

Hare, B., Call, J. & Tomasello, M. (2006) "Chimpanzees deceive a human by hiding." *Cognition*, 101, pp. 495–514.

Hauser, M., Cushman, F., Young, L., Jin, R. K. X., & Mikhail, J. (2007) "A dissociation between moral judgments and justifications", *Mind and Language*, *22*, pp. 1–21.

Krueger, J. (1998) "The bet on bias: A foregone conclusion?", *Psycoloquy*, 9, http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?9.46

Krueger, J., & Clement, R. W. (1994) "The truly false consensus effect: An ineradicable and egocentric bias in social perception", *Journal of Personality and Social Psychology*, 67, pp. 596–610.

Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980) "Sex stereotypes and social judgment", *Journal of Personality and Social Psychology*, 39, pp. 821–31.

Locksley, A., Hepburn, C., & Ortiz, V. (1982) "On the effects of social stereotypes on judgments of individuals: A comment on Grant & Holmes's 'The integration of implicit personality theory schemas and stereotypic images'", *Social Psychology Quarterly*, 45, pp. 270–73.

Malle, B. F. (2004) *How the Mind Explains Behavior: Folk Explanations, Meaning and Social Interaction* (Cambridge, MA: MIT Press).

Malle, B. F., Knobe, J., & Nelson, S. (2007) "Actor-observer asymmetries in explanations of behavior: New answers to an old question", *Journal of Personality and Social Psychology*, 93, pp. 491–514.

Marks, G., & Miller, N. (1987) "Ten years of research on the false-consensus effect: An empirical and theoretical review", *Psychological Bulletin*, 102, pp. 72–90.

Martin, C. L. (2000) "Cognitive theories of gender development", in: T. Eckes & H. M. Trautner (Eds.), *The Developmental Social Psychology of Gender*, pp. 91–121 (Mahwah, NJ: Erlbaum).

Matsuzawa, T. (2006) "Evolutionary origins of the human mother-infant relationship", in: T. Matsuzawa, M. Tomonaga, & M. Tanaka (Eds.), *Cognitive Development in Chimpanzees*, pp. 127–41 (Tokyo: Springer).

McGrew, W. C. (1992) *Chimpanzee Material Culture: Implications for Human Evolution* (Cambridge: Cambridge University Press).

McGrew, W. C. (2009) "Ten dispatches from the chimpanzee culture wars, plus revisiting the battlefronts", in: B. G. Galef & K. N. Laland (Eds.), *The Question of Animal Culture* (Cambridge, MA: Harvard University Press).

McGrew, W. C., & Tutin, C. E. G. (1978) "Evidence for a social custom in wild chimpanzees?", *Man*, 13, pp. 234–51.

Mitani, J. C., Watts, D. P., & Muller, M. N. (2002) "Recent developments in the study of wild chimpanzee behavior", *Evolutionary Anthropology*, 11, pp. 9–25.

Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., Vanderklok, M. (1985) "The false consensus effect: A meta-analysis of 115 hypothesis tests", *Journal of Experimental Social Psychology*, 21, pp. 262–83.

Onishi, K., & Baillargeon, R. (2005) "Do 15-month-old infants understand false beliefs?", *Science*, 308, pp. 255–58.

Parr, L. A. (2003) "The discrimination of facial expressions and their emotional content by chimpanzees (*Pan troglodytes*)", in: P. Ekman & J. J. Campos & R. J. Davidson & F. B. M. de Waal (Eds.), *Emotions Inside Out: 130 Years After Darwin's The Expression of the Emotions in Man and Animals*, 100, pp. 56–78 (New York: Annals of the New York Academy of Sciences).

Perry, S., Baker, M., Fedigan, L., Gros-Louis, J., Jack, K., MacKinnon, K. C., Manson, J. H., Panger, M., Pyle, K., & Rose, L. (2003) "Social conventions in wild white-faced capuchin monkeys", *Current Anthropology*, 44, pp. 241–58.

Phillips, A. T., Wellman, H., & Spelke, E. (2002) "Infants' ability to connect gaze and emotional expression to intentional action", *Cognition*, 85, pp. 53–78.

Rholes, W., & Ruble, D. N. (1984) "Children's impressions of other persons: The effects of temporal separation of behavioral information", *Child Development*, 57, pp. 872–78.

Ross, L., & Nisbett, R. E. (1991) *The Person and the Situation: Perspectives of Social Psychology* (Philadelphia, PA: Temple University Press).

Royzman, E. B., Cassidy, K. W., & Baron, J. (2003) "'I know, you know': Epistemic egocentrism in children and adults", *Review of General Psychology*, 7, pp. 38–65.

Ruffman, T. & Perner, J. (2005) "Do infants really understand false belief?", *Trends in Cognitive Science*, 9, pp. 462–63.

Russell, J. (1987) "Can we say . . . ? Children's understanding of intensionality", *Cognition*, 25, pp. 289–308.

Trevarthen, C. (1979) "Communication and co-operation in early infancy: A description of primary intersubjectivity", in: M. Bullowa (Ed.), *Before Speech*, pp. 321–47 (Cambridge: Cambridge University Press).

Van Schaik, C. P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C. D., Singleton, I., Suzuki, A., Utami, S. S., & Merrill, M. (2003) "Orangutan cultures and the evolution of material culture", *Science*, 299, pp. 102–05.

Warneken, F., & Tomasello, M. (2006) "Altruistic helping in infants and young Chimpanzees", *Science*, 311, pp. 1301–03.

Wellman, H. M. (1990) *The Child's Theory of Mind* (Cambridge, MA: MIT Press).

Wellman, H., Cross, D., & Watson, J. (2001) "Meta-analysis of theory-of-mind development: The truth about false belief", *Child Development*, 72, pp. 655–84.

Whiten, A. (2007) "Cultural panthropology", *Understanding Chimpanzees: The Mind of the Chimpanzee*. Lincoln Park Zoo, March 24, 2007.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E. G., Wrangham, R. W., & Boesch, C. (1999) “Cultures in chimpanzees”, *Nature*, 399, pp. 682–85.

Wilson, T. D., & LaFleur, S. J. (1995) “Knowing what you’ll do: Effects of analyzing reasons on self-prediction”, *Journal of Personality and Social Psychology*, 68, pp. 21–35.

Zahavi, A. (1977) “The cost of honesty (Further remarks on the handicap principle)”, *Journal of Theoretical Biology*, 67, pp. 603–05.