It's in your nature: a pluralistic folk psychology

Kristin Andrews

Abstract   I suggest a pluralistic account of folk psychology according to which not all predictions or explanations rely on the attribution of mental states, and not all intentional actions are explained by mental states. This view of folk psychology is supported by research in developmental and social psychology. It is well known that people use personality traits to predict behavior. I argue that trait attribution is not shorthand for mental state attributions, since traits are not identical to beliefs or desires, and an understanding of belief or desire is not necessary for using trait attributions. In addition, we sometimes predict and explain behavior through appeal to personality traits that the target wouldn't endorse, and so could not serve as the target's reasons. I conclude by suggesting that our folk psychology includes the notion that some behavior is explained by personality traits—who the person is—rather than by beliefs and desires—what the person thinks. Consequences of this view for the debate between simulation theory and theory theory, as well as the debate on chimpanzee theory of mind are discussed.

Keywords   Folk psychology · Action theory · Belief attribution · Explanation · Prediction · Animal cognition · Developmental psychology · Social psychology · Simulation theory · Theory theory · Chimpanzee mind · Theory of mind

1 Introduction

The everyday human practices of predicting, explaining, interpreting, judging, coordinating and otherwise socially interacting with others are grounded in an understanding

K. Andrews (✉)
Department of Philosophy, York University, 4700 Keele St.,
Toronto, ON, Canada M3J 1P3
e-mail: andrewsk@yorku.ca

of others as minded agents. This understanding of others is commonly understood as requiring the ability to attribute mental states—paradigmatically beliefs and desires. However, there is good reason to think that the commonsense conception of the mind is pluralistic, and not limited to the attribution of propositional attitudes. While it is true that part of the commonsense view is that people act for reasons, and that those reasons are mentally represented propositions, the folk also think that people act because of who they are as persons. I will argue that people predict and explain behavior by appeal to personality traits, and that at least some personality traits cannot be understood as an oblique reference to beliefs, desires, or any other propositional attitude.

My account is at odds with the traditional view of folk psychological understanding of other minds, which emphasizes the attribution of propositional attitudes (Churchland 1981). Though some recent work in folk psychology recognizes that our understanding of other minds is not exhausted by our understanding of other's beliefs and desires (Gallagher 2006; Goldman 2006; Hutto 2004), the paradigmatic presentation of folk psychology remains wedded to the propositional attitude model, as is demonstrated in the recent theories offered by Nichols and Stich (2003) and Godfrey-Smith (2005).

The idea that folk psychology must involve the attribution of propositional attitudes is also reflected in the way we teach it. For example, in William Lyons's recent philosophy of mind textbook, he describes folk psychology using the following example:

> I might predict Mary's future behavior by saying, "We cannot count on Mary's vote because she will almost certainly not turn up to the Faculty meeting, because she believes that Fred will be there and she just cannot stand Fred" (Lyons 2001, p. 118).

As the story goes, we attribute beliefs and desires to Mary (the belief that Fred will come to the meeting, the desire to avoid Fred) plus a covering law (e.g. people do what they desire to do, *ceterus paribus*) and conclude that she will not attend the meeting.

While clearly a feature of the theory theory account of folk psychology, and some hybrids (e.g. Nichols and Stich 2003), the attribution of propositions is also implicated in various versions of simulation theory. For example, Alvin Goldman has argued that the attribution of mental states is necessary to get a simulation started. When beginning any mental simulation, he suggests we assume that the agent is like us, that we share relevant psychological features. This assumption makes us accept explanations that seem natural to us and reject those that are less natural (Goldman 1989). Goldman even suggests that the simulator should assume shared basic likings and desires, unless there is reason to think otherwise. In order to make these assumptions, we must attribute propositional attitudes.

The claim that we must predict and explain behavior via the attribution of beliefs and desires can be challenged on both empirical and theoretical grounds. Attributing personality traits to people in order to predict and explain is an important part of our folk psychology, and it is distinct from ascribing content to people's mental states. I will examine the relationship between the attitudes and traits by discussing how both are used in predicting and explaining behavior. This analysis demonstrates that trait attribution and belief/desire attribution are distinct practices, and reflect different ways we have of understanding others. The many ways in which we understand

others and their actions is not something that has been adequately acknowledged in the folk psychology debate, and recognizing the plural nature of our folk psychology has important implications.

2 Prediction

2.1 Prediction and attribution of the attitudes

It is perhaps too obvious to state that belief/desire attribution is neither sufficient nor necessary for predicting behavior. I may know that Fred believes there is a tiger in front of him, and that he doesn't want to be eaten by the tiger, but from that I can't predict whether he will run, back away slowly, or freeze. Conversely, I can predict that my student will show up at 3 p.m. for a make-up exam, because he said he would. To make the later prediction I need not know what motivates the student to act; I need only appeal to a simple inductive rule that people generally do what they say they are going to do.

However, one might object that it is necessary to have some underlying understanding of the target's attitudes before selecting and applying a heuristic. For example, one might claim that in order for me to understand what my student is saying, I must attribute propositions to him because true communication requires that the audience recognize that the utterer has a particular mental state. In this case, in order to understand the meaning of the acoustic blasts emanating from the student, I must believe that my student believes that he will be in my office at 3 p.m. ready to take the exam. Thus, one cannot use a simple heuristic like "people generally do what they say they are going to do" without attributing a belief to the utterer.

While it may be true that communication between two people involves the recognition that the other is an intentional agent, it doesn't follow that for each particular utterance made the audience must consider the mental states associated with that particular utterance. Indeed, there are some people who communicate but who seem not to have the ability to attribute beliefs. Children younger than 3 years old do not use false belief contrastives in conversation (Bartsch and Wellman 1995), nor do they pass the false belief task (Wellman et al. 2001). These young children are not thought to have much understanding of belief, and there is reason to think that they still lack a robust understanding of belief at age four. Even after they pass the false belief task, it takes several more years before children gain an understanding of the referential opacity of belief (Apperly and Robinson 1998, 2003). Belief is special this way; plenty of other mental state concepts, including the desire concept, develop much earlier.

Yet, children do appear to understand very well what others mean by their utterances before they pass the false belief task, and hence are able to treat others as intentional agents. For example, children who fail the false belief task are still able to answer memory questions about the events of a story, thus demonstrating some understanding of the story and the utterer's words (Wimmer and Perner 1983; Wellman et al. 2001). By two years, children are talking about desires (Bretherton and Beeghley 1982) and they describe themselves and others using intentional mental state terms other than belief (Dunn et al. 1987). This suggests that young children do recognize that others

are minded, intentional agents, but that this knowledge is not dependent on the ability to attribute beliefs. The kinds of predictions infants and toddlers can make, then, are not the result of belief attribution.[1]

Without explicitly acknowledging that another has beliefs with specified content, a child can use the "people generally do what they say they are going to do" heuristic in order to predict that father will chase her when he says, "I'm going to chase you!" The child need only know that the utterer is the right sort of thing to apply the heuristic to in order to use it successfully to predict behavior. In order to make largely accurate predictions, the child must realize that her father, unlike the tape recorder that emits the same sounds, is an intentional agent. However, on this model, the child doesn't need to attribute to him a particular belief associated with his utterance. She doesn't do that because she can't, not until she gains a more robust understanding of belief. Yet she is still able to make the prediction.

It's worth noting, as well, that nonhuman animals are skilled at predicting behaviors of their conspecifics, predators, and prey, even though it is doubtful that most other species understand that others have beliefs. Given that young children and nonhuman animals are able to predict behaviors without considering the content of another's belief, it follows that being capable of attributing both a belief and a desire is not necessary for predicting some behavior.

2.2 Prediction and trait attribution

Since young children can predict behavior without attributing beliefs, there must be methods for making predictions of intentional behaviors that do not rely on considering a person's primary reason for acting. Social psychological research suggests that rather than relying primarily on belief and desire attribution, humans make predictions based on a number of biases, stereotypes, trait and mood attributions, normative inferences about what people in such a situation ought to do, inductive generalization over past behavior, etc. In what follows I will focus on trait attribution as a method of predicting behavior.

Personality traits are properties of a person that are taken to be stable, and they are often used to describe behavioral dispositions. Humans use personality traits in many different contexts, such as when they are describing others. In one study where students were asked to provide confidential descriptions of their classmates, 65% of the terms used were trait terms (Park 1986). For example, classmates were more often described using personality trait terms such as being kind or smart, and less often described using physical descriptors such as having pink hair. Traits are also used to make sense of people's behavior, to form impressions of others, to make both predictions and explanations, and to arrive at judgments (Miller 1984; Ross and Nisbett 1991; Winter and Uleman 1984). Because people understand traits as stable and

---

[1] The standard explanation of the performance of children who fail to attribute false beliefs is that they lack a concept of belief. Note that this interpretation of the result rests on a representational view of belief, according to which believing that P involves having an attitude toward a mental representation that P. Thus, this discussion does not apply to dispositional accounts of belief, according to which believing that P involves having some behavioral disposition pertaining to P.

constant dispositions of an individual, they are used to make predictions of behavior; for example, a generous person is expected to be generous both at work and at home, and barring any unforeseen events, she is expected to remain that way through time.[2]

While we often rely on personality traits to predict behavior, it has been argued that this method is of only limited success. Predictions based on personality traits at the expense of relevant situational features often turn out to be false, because people often underestimate the power situations have over behavior (Ross 1977). This tendency, called the fundamental attribution error, has us basing our predictions more on internal characteristics of a person, such as moods or traits, than on external factors of the situation. While we often rely on character traits in order to figure out what someone will do next, there is evidence that the individual's environment and history plays a significant causal role in behavior, and that character traits play a lesser role than people expect. Experiments such as Milgram's infamous obedience experiments, where subjects were ordered to shock people at ever-increasing levels of voltage (Milgram 1963), and Darley and Batson's Princeton Seminary study, which found that theology students' willingness to help someone in need was inversely correlated with how rushed they were (Darley and Batson 1973), demonstrate the important role the situation has on a person's behavior. Such studies suggest that seemingly insignificant events could affect people's behaviors in ways they had not anticipated.

Folk psychology as the attribution of the attitudes is widely described in the philosophical literature as having robust predictive power. However, the social psychological literature suggests that our abilities to make accurate predictions via trait attributions are limited. One might conclude, then, that predicting through trait attribution is a less accurate method than predicting through appeal to beliefs and desires. However, given other findings in social psychology, this would be a hasty conclusion.

When humans form some expectation about how a person will behave, whatever behavior subsequently occurs is usually interpreted in such a way that it is consistent with the prior expectation. This is one possible reason that humans think they are more accurate predictors than they really are (Olson et al. 1996). In addition, humans tend to notice things that correspond to our beliefs and expectations, and so we're more likely to remember successful predictions (Kunda 2002). Thus, even though we may be poor predictors of behavior in some situations, it is possible that we won't notice the failure because we either interpret the situation so that it corresponds to our expectation, or we forget about making the faulty prediction.

Another explanation comes from Ross and Nisbett, who argue that our trait-based predictions are largely accurate because most of our predictions are made in the same situation, broadly construed (Ross and Nisbett 1991). For example, if you form a trait description of Sue in the workplace, and you only interact with Sue in the workplace, it is likely your prediction will be largely accurate, since traits are stable within situations. It is only when you attempt to generalize to another situation, such as how Sue would be as a road-trip companion, that your prediction is likely to fail. Ross

---

[2] There is some evidence that the reliance on traits over the situation may be to some extent a cultural artifact of western culture, and that Chinese, Koreans, or Japanese may be less likely to commit the fundamental attribution error (Nisbett 2003).

and Nisbett also suggest that the predictor is often part of the situation she predicts, and what may appear to be stability across situations may in fact only be the stability of the person's behavior in the predictor's presence. For these reasons, many of our predictions of behavior are correct.

This isn't to suggest that we only use trait attribution and belief/desire attribution when we predict behavior; trait attribution is just one additional method used. Our quotidian predictions that are so ubiquitous that they disappear—the prediction that (at least some) students will show up to class, that the driver in front of me will stop at the stop sign, that the waiter will bring the double espresso I ordered—are often based on other heuristics. Predictions of people we don't know can't appeal directly to trait attributions, because in order to develop some notion of a person's trait one must have some information about the person; the better we know a person, the better we are at judging the frequency of her attitudes and traits (Judd et al. 1991). When predicting the behavior of unfamiliar people we may use stereotypes, consensus effect (thinking others are like us), or social norms when making judgments about what others will do next.

Such heuristics involve inductive generalizations about typical behaviors over groups or individuals, and given the finding that people would be more likely to make accurate predictions of their own behavior if they used the base rates for people in their situation, our best bet at making accurate predictions in an unfamiliar situation is to determine what others have done before in a similar situation (Vallone et al. 1990). In familiar situations, and with familiar people, using inductive generalization over the person's past behavior may be the best strategy.

The social psychological picture suggests pluralism in the ways we make predictions, since we need not appeal to beliefs and desires in order to predict behaviors when we can appeal to personality traits or inductive generalizations over past behavior. These are methods of prediction that might be available to those who do not have a robust understanding of beliefs.

However, one might object that folk psychology's traditional emphasis on belief/desire psychology can be made consistent with the social psychological story, depending on the relationship between traits and propositional attitudes. Perhaps when I appeal to Sue's generous nature to predict that she will contribute to hunger relief, I simply mean that Sue wants to help others and believes that she can fulfill that desire by contributing to hunger relief. Understanding traits in this fashion reconciles at least part of the social psychological picture with the traditional philosophical one, and thus evidence from social psychology about trait attribution does not undermine the claim that we predict behavior through the attribution of beliefs and desires. The apparent conflict arises from the different terminology used in the different disciplines.

I don't think this argument is compelling. To see the problems with the argument, let us examine the possible relationships between trait attribution and the attribution of propositional attitudes.

We can start by examining the strongest relation, and ask whether trait attribution is identical to belief/desire attribution. By looking at the developmental literature, it seems clear that this relation won't hold, because children are only able to use personality traits long after they start to use mental state attributions. When a child is able to talk about and respond to others' beliefs and desires at age 31/2, she is still unable

to use traits to predict behavior (Kalish 2002; Rholes et al. 1990). While preschool-aged children are able to draw inductive inferences about stable physical traits (Carey 1985; Gelman 1988), they are unwilling to draw similar inferences about the stability of personality traits (see Miller and Aloise 1989; Rholes et al. 1990 for reviews). For example, Rholes and Ruble showed videos of characters acting in easily categorized ways to 5–10-year olds. Even when the younger children successfully described the behavior using a trait term, they were not likely to use the character's past behavior or the trait to predict future behavior (Rholes and Ruble 1984). Not until a child is in elementary school does she start using trait attributions to form her predictions (Cain et al. 1997; Heller and Berndt 1981; Heyman and Gelman 1999; Yuill and Pearson 1998). This is at least three years after the child is first thought to understand that people have beliefs that cause behavior. Given the developmental delay between belief attribution and personality trait attribution, the two abilities cannot be identical. Further, this evidence shows that the ability to attribute beliefs cannot be sufficient for attributing traits.

Another possible relation between traits and propositional attitudes is that the ability to attribute beliefs is necessary for attributing traits. On this view, traits are simply shorthand for beliefs and desires, and so the development of trait attribution relies on prior understanding of belief and desire. This possible relationship reflects the developmental data just discussed.

There is, however, empirical evidence against this view as well. Children who are unable to attribute mental states can come to learn how to predict behavior by attributing personality traits. Social Stories Therapy is an intervention method designed to teach social skills to children with autism (Gray 2000). Children on the autistic spectrum are taught how to recognize salient behaviors, apply a character trait or mood term to the behavior, and then associate other behaviors with that label. After successful therapy, the child is able to make predictions using the trait identifier, but she doesn't gain a theory of mind or an understanding of beliefs and desires as parts of a folk theory of behavior. For example, a person with autism might be taught to associate a smile with the term 'happy', and the term 'happy' with a number of specific behaviors (e.g. hugging, laughing, etc.). The child with autism can be taught to describe a smiling person as happy, and given that trait predict that the person will engage in happy behaviors. Though one might be inclined to describe the autistic child's new knowledge as a theory, it is a theory of behavior, not a theory of mental states.

The normal understanding of traits may be analyzed in a similar fashion. Rather than referring to mental states, a trait attribution could refer to a class of behaviors, and hence a person with or without autism can predict behavior without appealing to mental states. Though normals understand that others have mental states, and these mental states may explain a person's behavior, normals could appeal to a similar theory of behavior to predict action without appeal to the attitudes. If there is a normal folk theory of behavior, it would involve connecting specific traits with other traits, and with behaviors. There is some evidence for this view, since the data suggests that humans understand which sets of traits are consistent, and which are not, and such models are used when we predict what someone with certain traits will do (see Ross and Nisbett 1991 for a review). I take these considerations as reason to think that

without good evidence to the contrary, trait attribution ought not be considered shorthand for attribution of belief, on any non-dispositional account of belief.

To more fully examine the relationship between propositional attitudes and traits we must look at folk psychological explanation as well as prediction. One who thinks that an explanation for behavior will always involve at least an implicit appeal to one's primary reasons for acting—a person's belief and pro-attitude as Davidson would have it (Davidson 1963)—may well think that a personality trait will fail to offer a satisfactory explanation. However, if we can predict behavior through appeal to someone's personality traits without knowing the person's primary reason for acting, perhaps we use traits to explain behavior as well. For now, I hope to have demonstrated that when predicting behavior, we do not always need to attribute propositional attitudes. It follows that we should not expect to find a single mechanism that subsumes all our folk psychological predictions, since they are the result of a number of different heuristics, at least one of which, trait attribution, is not reducible to the attribution of beliefs and desires.

3 Explanation

3.1 Explanation and attribution of the attitudes

Research in attribution theory suggests that while reason explanations in terms of an actor's belief and desire are the most common means of answering why questions about behavior, not all good explanations will take that form. In one study of the frequency of different kinds of explanations people offer for behavior, 16% of the explanations did not refer to reasons (Malle 2004). Malle categorized the explanations people provided into three different categories, reasons, causal history, and enabling features, and found that 61% of the explanations referred to reasons alone, and 23% of the explanations combined reasons and causal history. While our methods of explaining behavior include attributing beliefs, desires, etc. to an individual, they also take account of external factors of the situation, the individual's perceived character traits, past behavior, and even one's own response to the situation (Franzoi 2003). This research suggests that the mechanisms utilized in offering explanation, like those implicated in prediction, are plural.

According to Heider's seminal work in attribution theory, people offer explanations of behavior in order to fulfill two needs: to form a coherent view of the world, and to control the environment (Heider 1958). Whereas the philosophy of science literature emphasizes the logical structure of explanation, with beliefs and desires as the explanans, the social psychology literature suggests a coherence model of explanation. The theory of explanatory coherence developed by Thagard (1989) is a more recent development of Heider's insight that we incorporate a number of different methods when explaining behavior. Thagard and Kunda apply this model to folk psychological explanation in order to describe how we "make sense" of other people, which they think involves three different methods: categorization in terms of traits, behaviors, social groups, etc., causal attribution, and analogy. They argue that while we rely on these different methods, they alone are not enough, for we also need a

coherence-maximizing system to oversee these different mechanisms and organize them into explanations that make sense (Thagard and Kunda 1997).

Research in social psychology confirms the ideas of Heider, Thagard, and others that we are pluralistic when it comes to developing explanations for behavior, and that coherence of some variety is important in constructing them. Evidence suggests that we rely on coherent narrative when offering explanations of behavior. In a well-known study of juror decision-making, it was found that lawyers who present their arguments in narrative form are more successful than lawyers who present arguments in some other, non-chronological order (Pennington and Hastie 1992). Such studies suggest that we construct imaginary scenarios in order to make sense of a person's behavior, scenarios that are coherent with all we know about the actors. So for example, when a person's past behaviors (e.g. she always brings wine to the party) and her traits (she is very generous) cohere with current behavior, we understand that behavior; it makes sense to us.

The psychological phenomenon of cognitive dissonance also suggests that we expect a consistent picture of others and ourselves. Cognitive dissonance refers to the feeling of discomfort that arises when we act in a way we take to be inconsistent with our self-image. Humans have a tendency to confabulate explanations for behaviors that may be caused by effects such as the fundamental attribution error, especially when those actions are contrary to their self-image. Thus, subjects in Milgram's obedience experiments might make up a story about why they administered dangerous shocks rather than admit that they are capable of harming others. And while we tend to take credit for our successes, attributing them to aspects of our character or our behavior, when we fail we tend to explain those actions by reference to external causes (Campbell and Sedikides 1999).

The mechanisms underlying our ability to explain behavior are very important to us, given that discussing people's behavior is one of the primary occupations of a human being. The evolutionary psychologist Robin Dunbar suggests that we spend two-thirds of our conversational time talking about what people are doing and why they are doing it (Dunbar 1996). This trend is evident among humans at a very young age; according to one study, as early as 4 years 77% of children's conversations are about people (O'Neill 2004). Though the statistics may be surprising, the general point shouldn't be: humans are social creatures, and we are very interested in other people.

A satisfying explanation of behavior will take advantage of a number of different available methods, and those methods will depend on what we already know about the person whose behavior we're explaining, as well as what we don't know. For example, while an inductive explanation may satisfy us when given to explain someone we don't know well ("Why is he drinking champagne? Is he celebrating something?" "No, he just has it every day with lunch."), such responses are often not very satisfying when we want to know the deeper motivation. Indeed, we will often provide an explanation in terms of induction over the person's past behavior only because we don't have any further information.

3.2 Explanation and trait attribution

Given the preceding discussion, it is perhaps already clear that trait attribution can also be used to offer explanations of behavior, even though it does not constitute a description of the reasons or the goal motivating the actor's behavior. In many cases, such explanations are the best that we can do.

One might object to explaining behavior in terms of trait attribution. For example, Owen Flanagan has expressed concern about explanations that refer to traits because they do not include the causal story leading to the behavior being explained, but only place the act in a category of other acts (Flanagan 1984). He argues that explaining why Jack stole the candy by saying that he is a juvenile delinquent either describes the behavior (i.e. to steal candy is to be a juvenile delinquent) or it amounts to an inductive explanation (e.g. Jack always does stuff like that). Despite these concerns, in some cases explanations in terms of trait attribution will be satisfactory; if you know about something common to all juvenile delinquents, that trait attribution might help you to understand why the child behaved as he did. Knowing that Jack is a juvenile delinquent helps us decide how to respond to his behavior, for example. Trait attribution can serve as an explanation when it yields additional information.

Note too that attributing traits to people when explaining their behavior may have some implications about the kinds of reasons a person might have had (or might not have had). For example, by knowing that Jack is a juvenile delinquent, you can infer that his reason for stealing the candy probably wasn't to test the store's security system, for example.[3] The class of possible reasons can be limited once you attribute a trait, but a number of belief/desire sets would still be consistent with the behavior so described. Perhaps Jack believed that stealing the candy was the only way to gain nourishment for himself or his younger brother. Perhaps Jack desired to cause problems for the shop-keep. Both explanations are consistent with the trait explanation that Jack is a juvenile delinquent.

Take another example. If Sue snaps at a student during office hours, and the student comes to me asking what he did wrong, I might explain Sue's actions by saying that she is testy. That trait attribution can make sense of her behavior, if the student knows that testy people snap for reasons unrelated to the actions of the snap-ee. That's not to say that trait attributions will provide the whole story; for someone who wanted a fuller explanation, the question might then become "why is she so testy", or "how did he become a juvenile delinquent?" and the answer to those questions would take the form of a more robust narrative. Like inductive explanations, we will sometimes offer trait explanations when we can do no better. Other times it is clear that we do not want any more in the way of an explanation—for example, when one explains a hated politician's speech by saying "He's either an idiot, or evil, or both."

Explanations that make sense are those in which the details are either filled in for us, or those where we can fill in the blanks. Attribution of mental states, traits, and generalizations about someone's past actions, along with the causal story of events, can provide us with a coherent and satisfactory explanation. Sometimes we can generate

---

[3] Thanks to an anonymous reviewer for this journal for making this point.

the description on our own, and other times we require more information in order to feel that we understand. What counts as understanding for an individual will depend on that person's background knowledge and her reasons for asking the why-question. Explanations of intentional behavior, like predictions of behavior, involve more (or less) than the attribution of propositional attitudes. Not all psychological explanations are reason explanations.

While traits are not shorthand for beliefs and desires, perhaps trait attribution is still a type of mentalistic attribution. Some trait attributions imply facts about a person's emotional state, such as being happy or testy. Explanation in terms of a person's emotions is an explanation in terms of the individual's mental states.

Traits themselves, however, are plural. Not all traits refer to a person's emotions, and there is reason to think that there are some traits that cannot be identified with an individual's mental states. For one, as I argued in the last section, there is a double dissociation between understanding traits and understanding beliefs. Normal children come to understand personality traits only after they understand the mental states of desire and belief, while some children who can't attribute beliefs are able to use traits to predict behavior. When impaired children use trait attribution, they are taught to understand traits as dispositions to behave in certain ways. In fact, this is how traits are understood in the psychological literature—as shorthand for kinds of behavior, rather than anything like private states of mind. Unless we are going to return to a behaviorist method of individuating mental states, personality traits should be seen as distinct from any mental states, including beliefs and desires.

To further defend the distinction, note that when we attribute a personality trait to an actor we are not always going to be attributing something that the actor would claim for herself. Or worse, the actor might have specific beliefs that contradict the personality traits. For example, I can predict that Mary won't hold the elevator for me because she is self-centered, even though Mary doesn't believe that she is self-centered. Negative personality traits in particular can be useful for predicting and explaining behavior, even though they do not make up a mental state of the individual, and even though the individual might deny that the trait term correctly describes her behavior. If Mary were asked to explain why she didn't hold the elevator, she might say, "I didn't notice you." However, it was the lack of attention to her surroundings that earned her the trait of being self-centered to begin with. Or she may not have an answer at all; as Davidson points out, we can act intentionally yet not be able to explain why we acted (Davidson 1963). Traits that are societally prohibited, such as being racist or sexist, are often not claimed by people whose behavior makes them accurately described as such. Accompanying the racist behavior may be all the narratives of anti-racism, and the unaware racist could firmly believe on intellectual grounds in the equality of the races to such an extent that she joins anti-racist organizations and rallies for equality. However, at the same time she might unknowingly discriminate against members of the race when they are not part of her social network or class. If you asked her why she discriminated, she may deny that she acted in such a way, claim she doesn't know, or offer an alternative reason explanation for her action. An alternative reason explanation, however, cannot be accepted at face value. In this case, the actor is citing her reasons retrospectively, rather than before or during the action. She may be reconstructing the cause of her action just as she would for another actor.

Given the tendency to confabulate explanations after the fact, and the desire to present a positive self-image to an audience, an agent's own post hoc reason for her action can be inaccurate. Consider Ted Bundy's bizarre explanations for his crimes, or an anorexic's rationalization for her dieting. Sour grapes situations give rise to another case in which untrustworthy reason explanations are given by an actor—i.e. when the fox walks away from the grape vine muttering, "I didn't really want those sour grapes anyway." Maybe he is just trying to convince himself, but if he succeeds, he will have reconstructed his reason for acting after the fact.

If, as I've argued, traits cannot be either identical to or shorthand for sets of beliefs and desires, if they can conflict with one's beliefs and desires, and if they need not be something an actor claims for oneself, then there are good reasons to reject the idea that traits are a kind of mental state. Further, given the conflict that can exist between a person's mental states and her traits, we shouldn't expect trait attribution to correlate with some set of beliefs or desires an actor has either.

At this point, one might suspect that people are not explaining intentional behavior when they offer an explanation in terms of personality traits. Explanations in terms of traits, one might say, can only serve as explanations of unintentional action; all proper explanations of intentional behavior must be reason explanations.

There are at least two reasons for rejecting the claim that explanations in terms of traits can only explain unintentional action. For one, the objection doesn't conform to the ordinary way people think about intentional action, since at least in some cases we claim that people's actions are intentional even when the actor doesn't have a primary reason for those actions. This point is demonstrated by people's response to a story about a chairman of the board who doesn't care about the environment, but only cares about maximizing profits (Knobe 2003; Malle 2006). The folk take the chairman to be intentionally harming the environment when such harm is a known side effect of his business decision, even though he doesn't have any reason to harm the environment—he doesn't construe his own action qua harming the environment, but sees it qua maximizing profits. He only has reasons for the primary behavior, not the side effect. Like side effect cases, cases that involve acts of negligence, impolitesse, or bias may lead to judgments of intentional action where no reason explanation for the behavior is evident, because the actor doesn't construe her action in the same way as the observers do. That the actor doesn't construe her actions in that way may be explained in terms of her personality.

Further, as Hursthouse argues, some actions are arational—again, not done for reasons in the Davidsonian sense—but are still intentional because they wouldn't have been performed had the actor not been in the emotional state that she was (Hursthouse 1991). Like explanations in terms of personality traits, Hursthouse argues that explanations in terms of an actor's emotional state cannot be understood as shorthand for an explanation in terms of beliefs and desires, and that emotion explanations—or arational explanations in general—ought to be seen as an autonomous variety of psychological explanation. An example of arational action is a person's jumping up and down from excitement; we can take the jumping behavior to be intentional, and explain it by the person's excitement over some event, yet not be able to offer a traditional reason explanation for the behavior; we can't "indicate what it was about the action

that appealed" (Davidson 1963, p. 685), we don't know why she jumped rather than smiled, and the actor herself wouldn't be able to answer that question.

Hursthouse argues that the folk understand some intentional actions to be explained not by beliefs and desires, but by emotional mental states. I am suggesting that we ought to also recognize that the folk understand some intentional actions to be explained by nonmentalistic personality traits—some actions are explained by who we are rather than what we think.

Of course, appeal to traits will not always answer our why questions. In some cases, explanations in terms of traits can be seen as an oblique reference to a person's past history. This may be true even if the individual doesn't agree that a past experience affected her behavior. For example, a pious person may have come to be that way because of some behavior modification techniques used on him as a child, yet have no knowledge of this history. However, someone who is aware of the psychological effects of certain kinds of religious training may explain the pious behavior in terms of the person's early life experiences.

Trait attribution may come in many varieties. The kinds of traits I am concerned with here make up an interesting subset of trait attributions that resist being analyzed in terms of propositional attitudes. For example, in response to the question "Why did she want such a long engagement?" a perfectly normal response might be, "She's a careful person who takes things slowly." It isn't that she doesn't yet believe that her fiancée is the right person for her, nor is it that she desires more time to asses the relationship. She can be fully committed to her promise, and simply be the kind of person who doesn't tend to move quickly. In this way, the trait explanation is quite similar to Hursthouse's arational emotional explanation of behavior. Consider Hursthouse's example of an action explained by anger:

> Jane, who in a wave of hatred for Joan, tears at Joan's photo with her nails…I can agree that Jane does this because, hating Joan, she wants to scratch her face…I can agree that she would not have torn at the photo if she had not believed that it was a photo of Joan; and if someone wants to say, "So those are the reasons for the action," I do not want to quarrel, for these "reasons" do *not* form the appropriate desire-belief pair assumed by the standard account. On the standard account, if the explanatory desire in this case is the desire to scratch Joan's face, then the appropriate belief has to be something absurd, such as the belief that the photo of Joan *is* Joan, or that scratching the photo will be causally efficacious in defacing its original. And my disagreement is with adherents of the standard account, who must think that some *non*absurd candidates for appropriate beliefs to ascribe to agents performing arational actions are available. (Hursthouse 1991, pp. 59–60).

In both the engagement case and in the photograph case, there may be no appropriate set of beliefs and desires that constitute a reason for the action, and yet these actions are intentional. The trait and the emotion explain the behavior; Hursthouse claims that we consider the emotion to cause the behavior too. In like fashion, one might think that it is part of folk psychology that traits also cause action. For example, a father may attribute his reluctance to let his daughter date to being old-fashioned, recognizing that he has no rationale for his behavior. Being old-fashioned or conservative may be

construed as causing his decision, at least proximately.[4] This issue is not central to the argument I am defending here, but it is of interest given the significant departure from the standard view.

Given the folk's view that traits explain behavior, and that traits cannot always be identified with reasons, it is part of folk psychology that not all actions are explained by reasons. This view is at odds with the tradition. But it is a consequence of the plural nature of folk psychology. Human behavior, like the scorpion's stinging of the frog who kindly gave him a ride across the river, is sometimes explainable in terms of who we are, not what we think.

4 Conclusion

Attempts to uncover the cognitive architecture subsuming our folk psychological practices must take into account the plural nature of our understanding of other minds. Given the arguments in favor of the view that people use a variety of techniques when predicting and explaining behavior, we must reject the narrow definition of folk psychology as the attribution of belief and desire. We should accept a broader understanding of folk psychology, and give up putting forth monolithic theories meant to account for all instances of predicting and explaining behavior.

The plural nature of folk psychology has implications for many debates in philosophy and cognitive science. Let me briefly mention two.

First, there are consequences for the debates around the cognitive architecture subsuming folk psychology. Given that folk psychology includes more than the attribution of propositional attitudes, any theory of folk psychology that does not include a discussion of other methods of predicting and explaining behavior will be, at best, incomplete.

For example, simulation theory seems difficult to reconcile with folk psychology qua trait attribution. When we simulate in order to understand another, we replicate or emulate (Goldman 2006) another person's mental states. This might work well enough for beliefs, desires, emotions, and the like, but it is not obvious how simulation would permit others to identify a person's trait. Robert Gordon's version of simulation, in which a simulator becomes the target much like an actor becomes a character, might seem better suited to account for folk psychology that includes personality traits as well as mental states (Gordon 1995). However, Gordon's position takes the personality traits for granted; when you become another, you adopt their traits. Once you've identified with another person, perhaps you can predict and explain that person's behavior, but in order to simulate, one must first have an understanding of another's traits, and then use those traits in the simulation. How the traits are first identified and adopted remain outside of Gordon's account.

---

[4] There are important differences between causes in terms of emotions and causes in terms of traits. The experience of emotion is an event, and it raises no metaphysical concerns to consider the wave of hatred as causing the destruction. However, a personality trait, as a stable disposition of a person, is not an event. Traits, unlike beliefs, desires, emotional experiences, or other mental states cannot have the same sort of efficacy. If it is part of folk psychology that traits cause behavior, as well as explain behavior, then perhaps this metaphysical worry must be addressed.

Since one can learn about a person's traits discursively or through observation, a folk psychology that includes personality traits may be more compatible with the theory theory of folk psychology. Traits, along with mental states, could be theoretical entities that we use to predict and explain behavior.

A second implication of a plural folk psychology relates to empirical research on the ontogeny and phylogeny of theory of mind. Today, standard theory of mind tests utilize predictive paradigms. Subjects in the false belief task, for example, are asked to predict the behavior of a character with a false belief. Chimpanzees are asked to predict where a conspecific will go to obtain food. The methods for studying the understanding of other minds that rely on predictive tasks may not be the best way to determine whether an infant or an animal has that understanding, given the availability of other heuristics (Andrews 2005). We might expect to find more attribution of mental states in the explanation of behaviors. Researchers designing experiments to test for theory of mind in children and nonhumans are advised to look toward explanatory paradigms, rather than relying on the current set of predictive ones. The current research on chimpanzee theory of mind uses predictive paradigms in order to determine what sort of understanding chimpanzees have about others' minds. However, we shouldn't expect that chimpanzees use mental state attribution to predict behavior in many simple cases if humans don't. If researchers hope to find evidence that chimpanzees appeal to mental states to make simple predictions of behavior, they are hoping to find a more robustly mentalistic creature in the chimpanzee than we find in the human. Examining explanation-seeking behavior rather than predictive behavior will yield knowledge about how chimpanzees understand other minds—knowledge that has eluded us so far.

A more theoretical implication for animal cognition research has to do with the status of animals as social agents. If animals use induction over past behavior and trait attribution for making prediction, just as humans do as part of their folk psychology, then we should conclude that such animals have a form of folk psychology as well (Andrews 2007). Across species, animals need to predict the behavior of potential prey, predators, and mates, and in social species even more complex predictive behaviors are observed. Since most non-human animals probably do not have the ability to attribute complex propositional attitudes to others, the methods animals use to make predictions likely utilize non-mentalistic heuristics (like some trait attribution) and mentalistic methods (like emotion attribution) that are simpler than full-blown belief/desire psychology.

While there is no denying that having an understanding of beliefs and desires provides one with an advantage, the extent of that advantage has been overemphasized by those who assume that our folk psychology consists wholly of mental state attribution. If the accounts of prediction and explanation that I have offered here are correct, the consequence is that the mature human is not the only one with access to folk psychology. Rather, those aspects of folk psychology that rely on inductive generalizations and trait attribution are open to some of those who aren't able to attribute content to other minds. We may share aspects of our folk psychology with our non-human relatives, even if we are the only species to offer reason explanations.

References

Andrews, K. (2005). Chimpanzee theory of mind: Looking in all the wrong places? *Mind and Language, 20*, 521–536.

Andrews, K. (2007). Critter psychology: On the possibility of nonhuman animal folk psychology. In D. Hutto & M. Ratcliffe (Eds.), *Folk Psychology Reassessed*. New York: Springer.

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*, 297–311.

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

Bartsch, K., & Wellman, H. (1995). *Children talk about the mind*. New York: Oxford University Press.

Bretherton, I., & Beeghley, M. (1982). Talking about internal states. *Developmental Psychology, 18*, 906–921.

Cain, K. M., Heyman, G. D., & Walker, M. E. (1997). Preschoolers' ability to make dispositional predictions within and across domains. *Social Development, 6*, 53–75.

Campbell, W. K., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of General Psychology, 3*, 23–43.

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* (vol. 2, pp. 485–517). Hillsdale, NJ: Erlbaum.

Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*, 67–90.

Darley, J., & Batson, C. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology, 27*, 100–108.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy, 60*, 685–700.

Dunbar, R. (1996). *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.

Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Developmental Psychology, 23*, 132–139.

Flanagan, O. (1984). *Science of the mind*. Cambridge, MA: MIT Press.

Franzoi, S. L. (2003). *Social psychology*. New York: McGraw Hill.

Gallagher, S. (2006). The narrative alternative to theory of mind. In R. Menary (Ed.), *Radical enactivism: Intentionality, phenomenology, and narrative*. Amsterdam: John Benjamins.

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

Godfrey-Smith, P. (2005). Folk psychology as a model. *Philosophers' Imprint, 5*(6), 1–16.

Goldman, A. (1989). Interpretation psychologized. *Mind and Language, 7*, 161–185.

Goldman, A. (2006). *Simulating minds*. New York: Oxford University Press.

Gordon, R. (1995). Folk psychology as simulation. In M. Davies, & T. Stone (Eds.), *Folk psychology*. Oxford: Blackwell Publishers.

Gray, C. (2000). *Writing social stories with Carol Grey*. Arlington, TX: Future Horizons.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Heller, K. A., & Berndt, T. J. (1981). Developmental changes in the formation and organization of personality attributions. *Child Development, 52*, 683–691.

Heyman, G. D., & Gelman, S. A. (1999). The use of trait labels in making psychological inferences. *Child Development, 70*, 604–619.

Hursthouse, R. (1991). Arational actions. *Journal of Philosophy, 88*, 57–68.

Hutto, D. D. (2004). The limits of spectatorial folk psychology. *Mind and Language, 19*, 548–573.

Judd, C. M., Ryan, C. S., & Park, B. (1991). Accuracy in the judgment of in-group and out-group variability. *Journal of Personality and Social Psychology, 61*, 366–379.

Kalish, C. W. (2002). Children's predictions of consistency in people's actions. *Cognition, 84*, 237–265.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*, 190–193.

Kunda, Z. (2002). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.

Lyons, W. (2001). *Matters of the mind*. London: Routledge.

Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning and social interaction*. Cambridge, MA: MIT Press.

Malle, B. F. (2006). The relation between judgments of intentionality and morality. *Journal of Cognition and Culture, 6*, 61–86.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371–378.

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology, 46*, 961–978.

Miller, P. H., & Aloise, P. A. (1989). Young children's understanding of the psychological causes of behavior: A review. *Child Development, 60*, 257–285.

Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford, UK: Oxford University Press.

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently ... and why*. New York, NY: The Free Press.

Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies. In E. T. Higins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles*. New York, NY: Guilford Press.

O'Neill, D. (2004). Small talk and stories: Children's developing models of the self and others. Cultural Biology: Evolution, development, and mind conference. Waterloo, ON: University of Waterloo, May 21–22.

Park, B. (1986). A method for studying the development of impressions of real people. *Journal of Personality and Social Psychology, 51*, 907–917.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Personality and Social Psychology Bulletin, 16*, 90–105.

Rholes, W., Newman, L. S., & Ruble, D. N. (1990). Understanding self and other: Developmental and motivational aspects of perceiving persons in terms of invariant dispositions. In E. T. Higgins, & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (vol. 2, pp. 369–407). New York: Guilford Press.

Rholes, W., & Ruble, D. N. (1984). Children's impressions of other persons: The effects of temporal separation of behavioral information. *Child Development, 57*, 872–878.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (vol. 10, pp. 174–221). New York: Academic Press.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. Temple University Press.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brian Sciences, 12*, 435–467.

Thagard, P., & Kunda, Z. (1997). Making sense of people: Coherence mechanisms. In S. Read, & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior*. Hillsdale, NJ: Erlbaum.

Vallone, R. P., Griffin, D. W., Lin, S., & Ross, L. (1990). Overconfident prediction of future actions and outcomes by the self and others. *Journal of Personality and Social Psychology, 58*, 582–592.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Winter, L., & Uleman, J. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47*, 237–252.

Yuill, N., & Pearson, A. (1998). The development of bases for trait attribution: Children's understanding of traits as causal mechanisms based on desire. *Developmental Psychology, 34*(3), 574–586.