People's Folk Theory of Behavior

Bertram F. Malle
(bfmalle@darkwing.uoregon.edu)
Institute of Cognitive and Decision Sciences &
Department of Psychology
1227 University of Oregon
Eugene, OR 97403-1227

Abstract

The folk theory of behavior is a conceptual framework that guides all of people's dealings with behavior, including attention, explanation, and control. Philosophy of action and developmental research into children's "theory of mind" have relied heavily on plausible but speculative assumptions about this folk theory. The present paper describes empirical research on three key elements of the theory, as found in the adult social perceiver: (a) how people conceptualize intentionality and differentiate intentional from unintentional behavior; (b) which types of behavior (intentional vs. unintentional, observable vs. unobservable) they attend to and choose to explain; and (c) how they explain these behaviors.

People's folk theory of behavior comprises concepts and distinctions that hang together as a framework within which human behavior can be described and explained. As a conceptual framework, it does not guarantee true beliefs about specific behaviors; but it dictates how people fundamentally think about human behavior. As such, it guides and critically influences how people perceive, evaluate, and try to change behavior. The present paper describes empirical research into the key elements of this folk theory of behavior.

The Folk Concept of Intentionality

Intentionality is the core concept of people's folk theory of behavior. With this concept, people classify behaviors into two groups and treat them differently in attention, evaluation, explanation, prediction, and change. Many writers in philosophy and psychology have recognized the central role of intentionality, but they only speculated about the components of this concept. Malle and Knobe (1997a) explored empirically how people distinguish between intentional and unintentional behaviors. They demonstrated that people consider a behavior intentional if the actor has

- a *desire* for an outcome,
- *beliefs* about a behavior that leads to that outcome,
- a resulting *intention* to perform that behavior,
- the *skill* to perform the behavior, and
- *awareness* of performing it.

In a first study, the authors showed that people agree substantially in their judgments of intentionality for 20 verbally described behaviors (inter-rater agreement, α = 0.99). Most importantly, whether or not the instructions provided participants with an explicit definition of intentionality had no effect on average agreement, suggesting that intentionality is not just a theoretical construct but a folk concept that people spontaneously use to classify behavior.

In a second study people were asked to define directly what it means to "do something intentionally," and they reliably mentioned four components of intentionality: desire, belief, intention, and awareness. Study 3 demonstrated that when people make actual judgments of intentionality, they also require the presence of skill as a fifth necessary component of intentionality. They judged an agent to have flipped a penny to land on heads *intentionally* only if there was sufficient evidence for his acquired *skill* to do so. More important, people clearly distinguished between the agent's *intention* to make the penny land on heads and his doing it *intentionally* (cf. Bratman, 1987). Appropriate *beliefs* and *desires* are necessary conditions for inferring an agent's *intention* to act. Inferring that the agent acted *intentionally,* however, requires her *skill* to fulfill that intention.

To our knowledge, all previous (psychological or philosophical) models of intentionality omitted one or more of the five empirically derived components, thus falling short of reconstructing people's actual concept of intentionality. In particular, the distinction that people make between an agent's intention and an actions' intentionality has important implications for how people evaluate behavior. That is, evaluations such as credit and blame rely both on the agent's antecedent mental states and on the action's skill-ful performance. This distinction has been underappreciated in both psychological, philosophical, and legal accounts of responsibility and blame.

Which Behaviors People Explain

The folk theory of behavior specifies what kinds of behaviors exist and how they can be perceived, explained, and controlled (cf. D'Andrade, 1987). Malle and Knobe (1997b) proposed that people's folk theory of behavior entails, on the most basic level, a 2 × 2 classification of behavioral events, which results from crossing the concept of intentionality (intentional vs. unintentional) with the concept of observability (observable vs. unobservable), as depicted in Figure 1.

	Intentional	Unintentional
Observable	actions	mere behaviors
Unobservable	intentional thoughts	experiences

Figure 1. Classification of Behavioral Events

In five studies, Malle and Knobe (1997b) showed that people wonder about and explain these behavioral events in systematically different ways, depending on three factors: (a) whether they are in the "actor" role (perceiving their own behaviors) or in the "observer" role (perceiving other people's behaviors); (b) whether they merely wonder-why or actually explain why a behavior occurs; and (c) whether they explain the behavior to themselves ("private" explanation) or to a conversation partner ("communicative" explanation).

Malle and Knobe (1997b) assumed that explaining why an event occurs requires a wondering why. They further assumed that such wonderings-why occur if three conditions are met: The person wondering is (1) aware of the event, (2) is in a subjective state of nonunderstanding (i.e., does not have an explanation available), and (3) finds this nonunderstanding relevant so as to be motivated to remove it. Malle and Kobe then showed that such wonderings can be about any of the four behavioral events distinguished above (see Figure 1), but the three principles of wondering predict reliable actor-observer differences in the selection of *which* behaviors people wonder about. Actors know (or think they know) the reasons for their intentional behaviors, so they do not wonder why they occur because they are not in a state of nonunderstanding; instead, they wonder about their own unintentional behaviors, for which they are in a state of nonunderstanding. Observers, by contrast, wonder about other people's intentional behaviors because these are diagnostic and socially consequential, but they wonder less about unintentional behaviors because being in a state of nonunderstanding with respect to them is seldom relevant. Furthermore, because actors have more direct access to their unobservable behaviors (e.g., feelings, thoughts) than to their observable behaviors (e.g., facial expressions, body language), they wonder more about unobservable behaviors than about observable behaviors. By contrast, observers have more access to other people's observable behaviors than to their unobservable behaviors, so they wonder more about observable behaviors. Two studies that coded daily thought protocols of college students and three 20th century novels supported all of these predictions: Actors wondered more about unintentional than intentional behaviors (73% to 27%) and more about unobservable than observable behaviors (67% to 33%), whereas observers wondered more about intentional than unintentional behaviors (67% to 33%) and more about observable than unobservable behaviors (74% to 26%).

Folk explanations, it was suggested, can be either private or communicative. A private explanation is a person's answer to her own wondering why, so people should privately explain the same behaviors that they wonder about. Two studies that coded college students' memory protocols of past behavior explanations and seven personal diaries showed indeed that actors privately explained more unintentional than intentional behaviors (74% to 26%) and more unobservable than observable behaviors (70% to 30%), whereas observers privately explained more intentional than unintentional behaviors (65% to 35%) and more observable than unobservable behaviors (74% to 26%).

A communicative explanation, by contrast, is an answer to another person's wondering, and this other person is always in the observer role. So observers should still explain the same types of behavior to other observers as they explain to themselves, whereas actors now explain those types of behavior to observers that the *observers* typically wonder about, namely, intentional and observable behaviors. Two studies that coded memory protocols as well as transcripts of conversations confirmed these predictions: In communicative contexts both actors and observers explain more intentional than unintentional behaviors (62% to 38%) and more observable than unobservable behaviors (70% to 30%).

In an unpublished study, MacConnell (1996) found that these actor-observer asymmetries regarding the types of behaviors actors and observers *explain* also extend to the types of behaviors they pay *attention* to. Pairs of undergraduate students were asked to have a getting-acquainted conversation and afterwards filled out a questionnaire in which they reported about that interaction. Specifically, participants described the behavioral events "that were going on" for themselves (actor role) and their partner (observer role). These reports of behavioral events were then coded for the intentionality and observability of each event mentioned. The results showed that actors reported more unobservable events (M = 5.7) than observers did (M = 3.5) and also reported more unintentional events (M = 5.9) than observers did (M = 4.6), p's < .01. Conversely, observers reported more observable events (M = 4.6) than actors did (M = 3.4) and also reported more intentional events (M = 3.6) than actors did (M = 3.1), p's < .01. Control conditions ruled out that these differences are due merely to differential reporting biases: Some participants were encouraged to report about *experiences*, others to report about *behaviors*, but the actor-observer differences remained. It seems likely, then, that the reports of different behavioral events are due to differential *attention* to those events in the first place.

If actors and observers attend to and explain different behavioral events, we may hypothesize that certain interpersonal conflicts are difficult to resolve because the two partners are actually arguing about different behavioral events. For example, one person (the actor) may focus on her *experience* of being angry and hurt, whereas her partner (the observer) is focusing on the actor's overt *action* of yelling. The conflict may persist as long as the two people continue to talk about two different behavioral events.

How People Explain Behavior

We have seen that people distinguish between different types of behavior and that they explain some types more frequently than others (Malle & Knobe, 1997b). But do people explain all types of behavior the same way? According to the most widely used psychological theory of behavior explanations, "attribution theory" (e.g., Kelley, 1967), people indeed explain all behaviors with causes—

either person causes or situation causes. By contrast, many philosophers have argued that people explain human action with *reasons* (e.g., Audi, 1993; Dretske, 1988), which are mental states (mostly beliefs, desires) in light of which the agent *decides* to perform the action. This apparent contradiction vanishes once we realize that people indeed use both modes of explanation: They explain intentional behaviors with reference to the agent's subjective reasons and conscious decision making, whereas they explain unintentional behaviors unintentional behaviors with "mechanical" causes that neither involve the agent's subjective reasoning nor any decision making (and often not even awareness). Malle (1997) has provided a detailed analysis of these two modes of explanation in a theoretical model of folk explanations. That model lays out the conceptual and linguistic properties of folk behavior explanations and derives numerous predictions about the role of these explanations in social perception and social interaction.

Modes of Explanation

The empirical study of naturally occurring behavior explanations reveals two major modes of explanation: *causal explanations* for unintentional behaviors and *reason explanations* for intentional behaviors. In addition, a less frequent third type is occasionally used to explain the causal history of an actor's reasons to perform an intentional behavior. These *causal history of reason explanations* are themselves a subtype of causal explanations because reasons (beliefs, desires) are unintentional behavioral events that are explained by causes.

To test the basic assumption that a behavior's intentionality predicts its mode of explanation, Malle (1997) presented 20 verbally described behaviors to one sample of students and recorded their judgments of intentionality. The same 20 behaviors were presented to a new sample who explained each behavior. These explanations were then reliably categorized as causal, reason, or causal history of reason explanations by four coders. The behaviors' intentionality rated by one sample almost perfectly predicted the probability of causal or reason explanations provided by the second sample, $r\,(20) = .93$. (Causal history of reason explanations occurred in only 10% of all explanations for intentional behaviors.)

In a second study, Malle (1997) reversed this logic by selecting two behaviors whose intentionality was ambiguous ("Anne drove above the speed limit" and "Vince interrupted his mother") and presenting each behavior with either a causal or a reason explanation. Participants then rated each behavior's intentionality. The results showed that people considered behaviors explained by reasons intentional but regarded *the same* behaviors explained by causes as unintentional.

This relationship between intentionality and type of explanation may be unsurprising to most philosophers, but it falsifies the basic tenet of social psychology's attribution theory and calls many classic findings of this research tradition into question. For example, a classic paper by Jones and Nisbett (1972) described an asymmetry between actors' and observers' explanations of behavior: Actors tend to explain their own behaviors by situation factors, whereas observers tend to explain others' behaviors by (dispositional) person factors. This asymmetry, however, applies only to the *causal* explanation of unintentional behaviors (for reasons are always "person factors"). The limitation of this classic actor-observer asymmetry as well as the existence of other, previously overlooked asymmetries can be derived from the particular properties of reason explanations, as described next.

Properties of Reason Explanations and Their Implications

Because reason explanations depict the actor's active reasoning process, they invoke the social norm of rationality (e.g., Bratman, 1987; Lennon, 1990; Pettit, 1993). Therefore, if the norm of rationality is made salient, people (as actors) should alter their reason explanations for self-presentational purposes. For example, they might provide more reasons for each behavior (to explicate their reasoning processes) and especially more belief reasons (which provide the rational link between the desire and the behavior at issue). In addition, under rationality pressures people might more often claim intentionality for equivocal behaviors (which could be interpreted as either intentional or unintentional) because only intentional behaviors allow a display of rationality.

Second, because reason explanations refer to the actor's own reasons, an explainer must fully take the actor's subjectivity into account. Only those reasons can come to explain intentional actions of which the actor is (at least dimly) aware at the time of acting. Since actors are often presumed to know the reasons of their actions, whereas observers can only infer them, actors are granted a sort of sovereignty over their reasons. The default way of understanding other people's intentional actions is therefore to ask them for their reasons. Moreover, actors may use their sovereignty to construct strategic reason explanations. The schoolboy's explanation why he started a fight may not reflect his actual reasons but the reasons that are most likely to appease his teachers. People cannot, however, freely invent reason explanations, because their social community will often scrutinize their reported reasons. For example, observers use tests of consistency to examine reason explanations for their truthfulness and coherence (Gustafson, 1986, ch. 5). Because intentional actions are explained against a whole background of beliefs and desires, reason explanations must be consistent with that background. Specifically, people examine the reported reasons for their consistency with other facts about the actor, about the situation, and about the culture they live in. Suppose Ben asks Anne why she went to the kitchen, and she claims, "To get water for your plants." If Ben distrusts her explanation, he might ask, "Since when do you care for my plants?" (actor's consistency) or, "But the plants aren't dry!" (situational consistency), or, "At 4 o'clock in the morning?" (cultural consistency).

Third, because intentional behaviors are caused by intentions, which are based on the actor's reasons, people will typically change an actor's intentional behavior by changing her reasons—for example, by adding a desire that overrides the previous desire or by changing a belief that

allows him to achieve the relevant desire via a different course of action. Attempts to change others' intentional behaviors will thus take the form of commands, persuasions, or negotiations (in which the norms of rationality and consistency will be highlighted).

Fourth, on a more linguistic level, reason explanations have at least three parameters: (1) they can be beliefs or desires; (2) their propositional content (*what* is desired or *what* is believed) can mention either an aspect of the agent or of the situation (re-invoking attribution theory's classic person-situation dichotomy); and (3) the reasons can be linguistically marked as mental states or not. For example, in explaining why Anne waters her plants twice a day, we may cite a desire reason that is marked ("because *she wants* them to grow faster") or unmarked ("so they grow faster"). Similarly, we may cite a belief reason that is marked ("because *she thinks* they'll grow faster") or unmarked ("because they'll grow faster").

These linguistic features of reason explanations, too, show asymmetries between the actor and the observer role. For example, belief reasons typically concern specifics of the decision to act (e.g., perceived circumstances, anticipated outcomes, considered alternatives), which may often be known only to the actor. Desire reasons, by contrast, appear to be relatively easy to infer from an observed action—at least for anybody who knows the culture's folk psychology (Bruner, 1990). Observers, who typically have less specific information available and must resort to easily inferable explanations, should therefore provide more desire reasons than actors do. Indeed, initial analyses of a growing database of naturally occurring explanations shows that observers provide, on average, more desire reasons (and fewer belief reasons) than actors do. As argued above, this asymmetry may also be due to actors' concern for presenting themselves in a rational light. Future research must sort out these varying functions of belief vs. desire explanations.

To provide another example for linguistic effects, mental state markers can have powerful communicative functions. For one, mental state markers may allow an observer to distance himself from an actor's reasons by emphasizing the actor's subjectivity ("She thought it was Monday . . ."). This effect was tested in a further study reported in Malle (1997). Undergraduate students read a vignette in which Cliff asks Jerry at a party, "Why did your girlfriend refuse dessert?" Jerry responds with an explanation that either contains a mental state marker (e.g., "She thinks she's been gaining weight") or does not (e.g., "She's been gaining weight"). After reading the vignette, participants rated how happy Jerry was with his girlfriend's current weight and how much they liked Jerry. If Jerry's use of a mental state functions to distance himself from his girlfriend's belief (that she has been gaining weight), Jerry should be seen as happier with his girlfriend's weight when he uses such a marker than when he doesn't. Indeed, Jerry was seen as significantly happier with her weight when he used the marker ($M = 5.4$) than when he did not ($M = 2.6$), $F(1, 43) = 20.6$, $p < .0001$.

Implications for Cognitive Process Models

The presented model of people's folk explanations targets explanations qua private judgments and communicative actions. The model does not specify the cognitive processes that underlie those judgments and actions. However, the model puts constraints on process theories by identifying certain tasks that must be accomplished on the cognitive process level.

To explain a behavior, the human cognitive system must determine the behavior's intentionality. In some cases, that assessment is effortlessly made in the course of perceiving or identifying the behavior. This occurs probably with the help of perceptual cues (when observing a behavior) or semantic cues (when hearing a behavior description) that a process theory would need to specify. In other cases, the perceiver collects and considers several pieces of information before deciding whether the behavior was intentional (e.g., in a court case). Such deliberated judgments of intentionality are likely to consider the five components of intentionality (see Malle & Knobe, 1997a), and the perceived presence or absence of each component determines the final intentionality judgment. A process theory would need to describe how these components are activated and combined (e.g., through symbolic or connectionist operations) into a judgment of intentionality.

If the behavior is judged unintentional, then the perceiver will begin a search for causes of the behavior and sometimes mechanisms that connect the causes with the behavior. This search will be guided by such factors as information about current situational forces impinging on the actor, general knowledge about the kind of behavior observed, knowledge about the actor, covariation information, and knowledge about the audience to whom the explanation is given. This task of searching through many sources of information to identify a cause for the behavior, and perhaps even a mechanism, requires multiple category activation and information integration that would need to be described in connectionist, probabilistic, or other terms.

If the behavior is judged intentional, perceivers need not search for mechanisms because the intentionality mechanism is always the same (reasons lead to an intention, which causes the action in the presence of skill and awareness). Thus, people need only search for reasons. The conditions of subjectivity and rationality constrain this search considerably: The perceiver looks for beliefs and desires (a) that the agent held at the time of acting, (b) that would provide rational grounds for acting that way, and (c), if the explanation is communicative, that would be informative for the explainer's audience. Such reasons will be inferred from characteristics of the action itself, the contrast event ("She did this rather than that"), the context, and relevant cultural scripts. Most important, however, reasons will be inferred from knowledge about the agent's mental states. In fact, the constraints of subjectivity and culturally shared rationality may invite perceivers to simulate (through empathy and perspective taking) the agent's subjective reasoning process (Goldman, 1989; Gordon, 1992).

These cognitive activities need to be described and accounted for by process theories of causal reasoning. Extant psychological theories of causal reasoning (e.g.,

Cheng & Novick, 1990; Kelley, 1967; Kruglanski, 1989; Read, 1987; Hilton & Slugoski, 1986) seem to have two limitations: First, they leave out the process of judging intentionality (which occupies significant cognitive resources in everyday life); second, and more important, they treat all explanatory reasoning as a search for causal explanations. The above analysis should make clear that a search for reasons differs significantly from a search for causes—in the concepts that are activated (e.g., rationality), the information considered (e.g., the agent's subjective mental states), and the cognitive heuristics used (e.g., mental simulation).

Domain-general or domain-specific? The previous implications suggest that process theories of explanatory reasoning may not be uniform and general because the processes of causal search and reason search differ significantly. Along these lines, recent work in developmental psychology has generated domain-specific models that deal with the tasks of perception, explanation, and prediction of human behavior (e.g., Baron-Cohen, 1995; Hirschfeld & Gelman, 1994; Leslie, 1995; Premack & Premack, 1995; Wellman, 1990). Repeatedly, the case of autism has been cited to support this domain-specific view since autistic children lack the folk conceptual framework for reasoning about mental states but have average or above-average capabilities in causal reasoning about physical events (Baron-Cohen, Leslie, Frith, 1985, 1986; for reviews see Baron-Cohen, 1995; Leslie, 1992). Even though there is considerable consensus that this "theory of mind" module exists on the functional-cognitive level, however, many scholars remain skeptical about the evidence for a corresponding specific brain substrate (e.g., Baldwin & Moses, 1995).

The theory of mind module is typically contrasted with a module that deals with "mechanical" causality in the natural world (see Carey, 1995), which some argue is even derivative of the social one (Premack & Premack, 1995). The question arises, however, whether unintentional behavior is processed by the theory of mind module or by the mechanical module; or might there be three modules?

Criticizing this proliferation of domain-specific modules, some scholars have postulated a domain-general causal reasoning framework that has a number of specialized and partially modified applications (see Keil, 1995, for a discussion). Sperber (1994) offered a compromise that appears particularly attractive. In his discussion of different levels of mental processes, he distinguishes between first-order modules, which are domain-specific, and second-order (metarepresentational) modules, which are domain-general. Applying this distinction to the problem of explanations, we may speculate that reason explanations of behavior are processed mainly by a first-order module for intentional behavior, whereas causal explanations are processed by a first-order module for unintentional behavior. The two modules may be activated, respectively, by an intentionality detector (cf. Baron-Cohen & Ring, 1994; Premack & Premack, 1995). These first-order modules interact, then, with a general causality processor (that also deals with "mechanical" causality) and perhaps a general hypothesis tester (cf. Kruglanski, 1989). These second-order modules would constrain first-order processes by enforcing general principles such as temporal order (causes must occur before effects), probability relations, and logical consistency.

To conclude, the present analysis of intentionality and folk explanations of behavior entails several process implications that new causal reasoning models need to account for. The question whether explanatory reasoning about behavior is processed in a domain-general or in separate domain-specific modules cannot be answered to date but is likely to prompt interesting future research.

Folk explanations and science A final point of discussion is the relation between folk explanations and scientific explanations of behavior. Here a common misunderstanding must be rectified. The adequacy of a social-psychological theory of folk explanations, such as the one presented, is independent of folk psychology's ultimate validity (cf. Goldman, 1993). Whether or not folk psychology is flawed does not detract from the fact that it guides social perception and behavior. Folk explanations of behavior are social phenomena that are open to scientific study just as other social phenomena are. A model of folk explanations is thus successful if it describes, explains, and predicts people's actual folk explanations. Whether those folk explanations are "objectively" valid as accounts of human action and whether they are founded on valid assumptions about the ontology of the human mind are entirely different questions (cf. Christensen & Turner, 1993).

In the same vein, research on the objective phenomenon of intentionality (e.g., G. Goldberg, 1985; Libet, 1985) and on the distinction between automatic and controlled processes (Posner & Snyder, 1975; Schneider & Shiffrin, 1975; Bargh, 1994) neither falsifies nor verifies research on the folk concept of intentionality. The two approaches are largely orthogonal: The phenomena under study are, in the one approach, objective processes of the mind and, in the other approach, people's theories about the mind. Of course, people's acts of reasoning within their folk theories are themselves cognitive processes, and they can be elucidated by objective process theories of the mind. But these cognitive theories describe the mind that reasons, not the mind reasoned about. In the end, a convergence between the objective nature of the mind and people's folk theories about the mind should come as no surprise since folk psychology presumably arose from people's dealings with the objective regularities of other minds—that is, from social interactions with other intentional beings (Brothers, 1985; Graham, 1987; Humphrey, 1984).

References

Audi, R. (1993). *Action, intention, and reason.* Ithaca, NY: Cornell University Press.

Baldwin, D. A., & Moses, L. J. (1994). The mindreading engine: Evaluating the evidence for modularity. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 13*, 553-560.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of Social Cognition* (2nd ed., Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Mechanical, behavioral, and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4*, 113-125.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37-46.

Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience, 1*, 27-51.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Carey, S. (1995). On the origin of causal understanding. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary symposium*. Oxford: Clarendon.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality & Social Psychology, 58*, 545-567.

Christensen, S. M., & Turner, D. R. (Eds.). (1993). *Folk psychology and the philosophy of mind*. Hillsdale, NJ: Erlbaum Associates.

D'Andrade, R. (1987). A folk model of the mind. In D. Holand & N. Quinn (Eds.), *Cultural models in language and thought*. New York: Cambridge University Press.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: The MIT Press.

Goldberg, G. (1985). Supplementary motor area structure and function: Review and hypotheses. *The Behavioral and Brain Sciences, 8*, 567-616.

Goldman, A. I. (1989). Interpretation psychologized. *Mind and Language, 4*, 161-185.

Goldman, A. I. (1993). The psychology of folk psychology. *Behavioral and Brain Sciences, 16*, 15-28.

Gordon, R. M. (1992). The simulation theory: Objections and misconceptions. *Mind and Language, 7*, 11-34.

Graham, G. (1987). The origins of folk psychology. *Inquiry, 30*, 357-379.

Gustafson, D. F. (1986). *Intention and agency*. Dordrecht: Reidel.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review, 93*, 75-88.

Hirschfeld, L. A., & Gelman, S. A. (1994). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Humphrey, N. (1984). *Consciousness regained*. Oxford: Oxford University Press.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.

Keil, F. C. (1995). The growth of causal understandings of natural kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary symposium*. Oxford: Clarendon.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press.

Kruglanski, A. H. (1989). *Lay epistemics and human knowledge*. New York: Plenum Press.

Lennon, K. (1990). *Explaining human action*. La Salle, IL: Open Court.

Leslie, A. M. (1992). Autism and the 'Theory of mind' module. *Current Directions in Psychological Science, 1*, 18-21.

Leslie, A. M. (1995). A theory of agency. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary symposium*. Oxford: Clarendon.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8*, 529-566.

Mac Connell, K. L. (1996). *Actors and Observers Attend to Different Behaviors*. Unpublished Senior Honors Thesis. University of Oregon, Eugene, Oregon.

Malle, B. F., & Knobe, J. (1997a). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*, 101-121.

Malle, B. F., & Knobe, J. (1997b). Which behaviors do people explain? A basic actor-observer asymmetry. *Journal of Personality and Social Psychology 72*, 288-304.

Malle, B. F. (1997). *Intentionality in attribution: A theory of people's folk explanations of behavior*. Manuscript under review.

Pettit, P. (1993). *The common mind: An essay on psychology, society, and politics*. New York: Oxford University Press.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55-85). Hillsdale, NJ: Lawrence Erlbaum Associates.

Premack, D., & Premack, A. J. (1995). Intention as psychological cause. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate*. New York: Clarendon.

Read, S. J. (1987). Constructing causal scenarios: A knowledge structure approach to causal reasoning. *Journal of Personality & Social Psychology*, 52, 288-302.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld, & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Wellman, H. (1990). *Children's theories of mind*. Cambridge, Mass.: MIT Press.