

8

Self-Other Asymmetries in Behavior Explanations Myth and Reality

BERTRAM F. MALLE
University of Oregon

A moment's thought suggests that people explain their own behavior quite differently from the way they explain other people's behavior. People know more (or different things) about their own behavior, they have different goals when explaining it, and they probably express their explanations differently in language. That the self differs from others as a target of explanation is a straightforward hypothesis, seemingly an obvious truth. But this chapter describes how the straightforward becomes complicated and the obvious unforeseen when we carefully examine a psychological phenomenon rather than take at face value what is generally claimed in textbook psychology.

So explanations of behavior performed by self and other differ. But exactly how? And why?

There are classic answers to this pair of questions (Jones, 1976; Jones & Nisbett, 1972; Watson, 1982). First, explanations for self and other differ in terms of dispositional and situation causes—people explain, in the “actor” role, their own behavior using situation causes and, in the “observer” role, other people's behavior using dispositions. Second, this classic “actor-observer asymmetry” holds because people know more about their own experiences, intentions, and history and because, as actors, they attend to the situation and, as observers, they attend to the other person's behavior (Jones & Nisbett, 1972).

In this chapter I examine the viability of these answers, both theoretically and in light of the existing evidence, and then discuss a pair of alternative answers. This alternative, to give a brief preview, suggests that explanations for self and other do not diverge in the use of person and situation causes but in other, more important aspects of explanation—for example, whether the explanation concerns intentional or unintentional behavior and whether it refers to reasons or causal histories, beliefs or desires. Furthermore, I will suggest that these self-other asymmetries exist

primarily because people explain their own versus another person's behavior on the basis of different information and with different pragmatic goals in mind.

SELF-OTHER ASYMMETRIES: A MAP

Before examining self-other asymmetries in behavior explanations, I should locate these specific kinds of asymmetries within a map of the broader landscape of self- and other-cognition.

Many comparisons have been made between self and other as the two major targets of cognition. Research has shown generally better memory for self-related information (Rogers, Kuiper, & Kirker, 1977), greater enhancement in self-descriptions and evaluations (Greenwald, 1980; Locke, 2002; Taylor & Brown, 1988), and considerable influence of own preferences and knowledge on the predictions about others' preferences and knowledge (Krueger, 1998; L. Ross, Greene, & House, 1977; Nickerson, 1999). Moreover, attention in social interaction is unevenly distributed such that people (as actors) routinely attend more to their own mental states and (as observers) more to the other person's observable actions (Malle & Pearce, 2001; Sheldon & Johnson, 1993).

All these patterns reflect differences in general processes such as memory, attention, evaluation, and judgment. Explanations of behavior should be considered a category in their own right, because they serve unique purposes, even though they rely on each of these general processes. In particular, behavior explanations have two main functions in social life: They are a cognitive tool with which people create meaning, and they are a social tool with which they manage social interactions (Malle, 2004). This dual function subjects behavior explanations to a variety of processes that can generate self-other asymmetries, such as knowledge structures, attention, and impression management. Self-other asymmetries in behavior explanation thus reflect important cognitive and motivational differences in self- and other-cognition but implement these differences in unique ways, owing to the specific social functions that behavior explanations serve.

An important landmark on the map of self-other cognition and explanation is the distinction between *which* behaviors actors and observers explain and *how* they explain them. If we don't assume with Kelley (1967) that all explained events are essentially the same, we can ask what *types* of events people are interested in explaining, and what we find are robust self-other differences (Malle & Knobe, 1997b). People try to explain more of their own *unintentional* than intentional events and more of their own *unobservable* than observable events. Events that are both unintentional and unobservable (which we can label *experiences*) are thus the single most frequently explained event type for oneself. By contrast, people try to explain more of others' *intentional* than unintentional events and more of their *observable* than unobservable events. Events that are both intentional and observable (which we can label *actions*) are thus the single most frequently explained event

type for others. To the extent that different event types elicit different kinds of explanations, self-other asymmetries will also emerge in the *kinds* of explanations used. Such explanation patterns triggered by differing event choices are intriguing and may even account for some of the classic actor-observer findings, but they are a topic treated elsewhere (Malle, 2004; Malle & Knobe, 1997b).

The focus of this chapter is on a different question: that of *how* actors' and observers' explanations differ for the same event—such as a particular action, a particular experience. I first explore the classic answer to this question and then propose a new approach.

THE CLASSIC POSITION: JONES AND NISBETT (1972)

In a famous paper, Jones and Nisbett (1972) formulated the hypothesis that “actors tend to attribute the causes of their behavior to stimuli inherent in the situation, while observers tend to attribute behavior to stable dispositions of the actor” (p. 93). For example, a senator might explain her vote against going to war by saying, “Preemptive wars are unjustified” whereas a political observer might explain the senator's vote by saying, “She is a soft-hearted liberal.”

Three features of the Jones-Nisbett thesis are noteworthy. First, the thesis is meant to apply to all behaviors—whether intentional or unintentional, observable or unobservable, positive or negative. Second, the thesis assumes that relevant attributions are of only two types: situational or dispositional. Third, however, attributions to *dispositions* turn out to be ambiguous (M. Ross & Fletcher, 1985), as they can refer either to any factor that lies within the person (including emotions, traits, beliefs, sensations) or specifically to stable traits. Though most discussions of the Jones-Nisbett thesis focused on the latter, the stable-trait meaning, strictly speaking we should consider two classic actor-observer asymmetries—one regarding person attributions in general, the other regarding stable trait attributions in particular. I return to this ambiguity later.

These are then the parameters of the classic actor-observer thesis: Attributions of any behavior can refer to either the situation or the actor's dispositions, and actors tend to explain their behavior by referring to the situation whereas observers tend to explain the same behavior by referring to the actor's dispositions.

What is the Evidence for the Classic Actor-Observer Thesis?

In the secondary literature on attribution research, the classic actor-observer asymmetry has been described as “firmly established” (Watson, 1982, p. 698), “robust and quite general” (Jones, 1976, p. 304), and “an entrenched part of scientific psychology” (Robins et al., 1996, p. 376). We learn that “evidence for the actor-observer effect is plentiful” (Fiske & Taylor, 1991, p. 73) and that there is “considerable evidence that the actor-observer bias is pervasive” (Aronson, 2002, p.

168). But there is, in fact, no thorough review available of the 30 years of research into the actor-observer asymmetry.¹ I therefore decided to conduct a meta-analysis on the primary literature of published actor-observer studies (Malle, 2005).

To survey the research literature on the actor-observer asymmetry in behavior explanation, I searched two databases (PsycINFO and Web of Science) for articles since 1973 that either contained relevant keywords (*attribution, actor, observer*) or cited the Jones and Nisbett (1972) paper. From this pool of about 1,500 articles, I selected those that (a) reported empirical data, (b) assessed explanations (“causal attributions”) instead of mere trait inferences from behavior, and (c) measured both genuine actor and genuine observer explanations (excluding studies, for example, that compare observer explanations with or without empathy instructions). Foreign-language articles were acceptable as long as I was able to conduct or obtain a translation into English. I then studied the results sections for reported data from which effect sizes could be computed (i.e., means, standard deviations, *F* or *t* values), in particular effect sizes for three dependent measures: (a) an overall difference score of person/disposition minus situation, (b) a separate person/disposition score, and (c) a separate situation score. In addition, where available, overall scores were computed separately for negative and positive events/behaviors.² Ninety-five articles with a total of 137 studies and 12,000 participants were examined.

As the effect size measure for each study, Cohen’s *d* was chosen, corrected for sampling bias following Hunter and Schmidt (1990, p. 281). The effect size averages in Table 8.1 were computed as weighted means, adjusted for differences in sample size across studies. The results are remarkable. Across all studies, we see a small overall effect of $d = 0.073$ (Unweighted means show the same pattern of results, with slightly higher numbers.)

These results hardly support the Jones-Nisbett hypothesis. For one thing, the overall effect is vanishingly small, corresponding to a correlation coefficient of less than .04. More important, the asymmetry obtains only for negative events ($r = .11$) and reverses for positive events ($r = -.13$). This pattern is consistent with a mild self-serving bias according to which actors, compared to observers, are more reluctant to explain negative events by reference to their own characteristics. The general claim, however, that actors explain their behavior with external causes and observers explain it in internal causes is not supported by 30 years of research.

TABLE 8.1 Meta-Analysis Results of the Traditional Actor-Observer Asymmetry for Causal Attributions, Based on 137 Studies

	Average Effect Size
Overall score ^a	0.073
Internal score only	0.045
External score only	0.000
Negative events/outcomes only (overall score)	0.221
Positive events/outcomes only (overall score)	-0.258

^aInternal-external difference score or interaction term for perspective (actor, observer) by locus (internal, external)

At this point we should take note that the hypothesized actor-observer asymmetry is neither pervasive nor robust, and certainly not firmly established, as textbooks of psychology would make us believe. But we need not necessarily conclude that actor-observer differences in behavior explanations are a myth; reality may just be more complex than past studies were able to reveal. One reason for the highly fluctuating results in testing the classic thesis may be that all studies relied on the dichotomy between personal/dispositional and situational attributions. This theoretical distinction has several drawbacks. First, the two categories are extremely vague in that the “person” category could be taken to indicate a behavior’s intentionality, the agent’s moral responsibility, or the agent’s conscious reason for acting, to name just a few. Likewise, the “situation” category could be taken to indicate a behavior’s unintentionality, the agent’s impunity, or an actual cause in the environment. Second, the person-situation dichotomy oversimplifies the way people explain behavior, forcing them to translate their potentially complex thoughts about the event in question into a crude bipolar language. It should be no wonder that enormous variations occur in such a translation, resulting in substantial error variance across studies. Perhaps actor-observer asymmetries in explanation really do exist, but to document them we need to do justice to people’s own concepts of behavior and explanation and to let them express those concepts in unfettered, natural ways. This would require giving up the simple person-situation dichotomy, and this is the approach I propose.

AN ALTERNATIVE APPROACH

Theoretical Framework

When people interpret human behavior, they sharply distinguish between intentional and unintentional behavior (Malle & Knobe, 1997a). The intentionality concept begins to form in infancy and helps the child perceive order and structure in the streams of human behavior (Baldwin, Baird, Saylor, & Clark, 2001), and it is a basic building block for human communication (Bretherton, 1991). With progressing age, the intentionality concept becomes more refined and is used to make moral judgments and plays a key role in promises, reciprocity, and other social contracts (e.g., Greenberg & Frisch, 1972; Kugelmass & Breznitz, 1968; Malle & Nelson, 2003). What is so important about the distinction of intentionality for behavior explanations is that people use different modes of explanation for each type of behavior (Malle, 1999, 2001, 2004).

As shown in Figure 8.1, when explaining *unintentional* behavioral events people use only one mode: that of causes. Cause explanations depict the factors that “mechanically” brought about the unintentional event—that is, without the agent’s control and often without the agent’s awareness. Consider these examples of cause explanations:

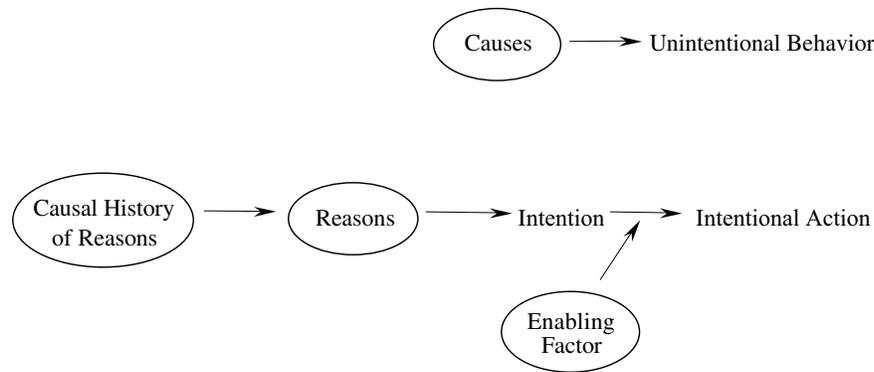


FIGURE 8.1

I almost failed my exams. *Why?* Oh, 'cause I didn't really prepare for them.
 A friend cried on the phone. *Why?* She felt that no one loved her.

When explaining *intentional* behavioral events, by contrast, people have to make a number of choices. The primary mode of explanation refers to the *reasons* the agent had for acting (Buss, 1978; Davidson, 1963; Locke & Pennington, 1978; Malle, 1999; Read, 1987). Reasons are seen as mental states (desires, beliefs, valuing) that the agent combines in a (sometimes rudimentary) process of reasoning, which then leads up to an intention and, if all goes well, to the intended action. So when people give reason explanations they really try to cite what was on the agent's mind when deciding to act—the reasons *in light of which* and *on the grounds of which* the agent acted. Suppose someone asks, “Why did Gina work 70 hours last week?” The conversation partner's explanation will likely cite one (or several) of Gina's reasons, such as “she wanted to impress her new boss,” “to get overtime pay,” “she knew that the project was due,” or “she was going on vacation the week after.” In all these explanations the explainer assumes two things: that the agent was subjectively aware of her reasons (she acted in light of them) and that these reasons rationally favored her course of action (she acted on the grounds of them). These two assumptions of *subjectivity* and *rationality* are the defining characteristics of reason explanations and differentiate them from all other modes of explanation. (For further discussion and evidence, see Malle, 1999; Malle et al., 2000; Mele, 1992; Searle, 1983.)

Under some conditions people cannot or do not want to offer reason explanations and therefore use one of two alternate modes of explaining intentional behavior. The first alternate mode refers to the *causal history of reasons*, which are factors that lie in the background of reason and clarify how these reasons came about. Consider again Gina's working 70 hours a week. If the explainer wanted to offer a causal history explanation, he might say, “she is driven to achieve,” or

“that’s the company norm.” Note that in the first of these causal history explanations, a disposition of the agent is mentioned (driven to achieve), whereas in the second, a situational factor is mentioned (company norm). It is important to keep in mind that the nature of causal history explanations has nothing to do with the person-situation dichotomy. Some causal history factors lie in the agent, others lie in the situation. What makes them causal history explanations is not their “causal locus” but the fact that they explain what brought about an intentional action and that they do so by referring to factors preceding the agent’s reasons. Because they precede reasons but are not themselves reasons, causal history factors neither meet the subjectivity assumption (the agent does not act in light of the factor) nor the rationality assumption (the agent does not regard the factor as grounds for acting).

To illustrate the distinction between reasons and causal histories, Table 8.2 contains two behaviors with contrasting reason explanations and causal history of reason (CHR) explanations.

The second alternate mode of explaining intentional action refers to factors that enabled the action to succeed. What these *enabling factor explanations* clarify is not what motivated the agent to act (which is explained by reasons or causal histories); rather, they clarify *how it was possible* that the agent’s intention turned into a successful action (McClure & Hilton, 1997; Turnbull, 1986). This “How possible?” question comes up for achievements (such as in academics, arts, or sports) but rarely for social actions (Malle et al., 2000), and as a result, enabling factors do not figure prominently in tests of actor-observer asymmetries.

Predictions

On the basis of this theoretical framework, what kinds of actor-observer asymmetries should we expect? Because, as previously mentioned, actors are generally concerned with explaining their own unintentional behaviors whereas observers

TABLE 8.2 Causal History of Reason (CHR) Explanations and Reason Explanations for the Same Behaviors

Behavior	Reason explanation	CHR explanation
<i>Kim chose not to vote in the last election.</i>	None of the candidates appealed to her.	She is lazy.
	She didn’t want to support the system.	She doesn’t realize that every vote counts.
<i>Brian used heavy drugs last Sunday at the party.</i>	He was curious what it would feel like.	He is a junkie.
	He thought it would be cool.	He grew up in a drug-dealing home.

Note. From Malle (1999, p. 35, Table 2).

are more concerned with explaining others' intentional behaviors (Malle & Knobe, 1997b), it follows that actors use more causes than observers do, since causes are the only mode of explanation for unintentional behaviors. This effect, however, is solely due to a choice of *which* behaviors one explains rather than a choice of *how* one explains them, that is, with what kinds of explanations. Far more interesting is the possibility that, for example, when explaining the same intentional behavior, actors use more reasons (and fewer causal histories) than observers do. This pattern constitutes a first actor-observer hypothesis, the reason asymmetry.

Hypothesis 1: Reason Asymmetry

Reasons are the default explanation mode of intentional action. They refer to the actual mental states that figured in the agent's reasoning that underlies the intention to act. Explainers—both actors and observers—will normally strive to offer reason explanations. In fact, on average, reasons make up about 70 percent of intentional action explanations. However, an explainer may offer a CHR explanation when either (1) information access is limited or (2) pragmatic goals favor causal history factors over reasons. A case of limited information access occurs when the explainer does not know the agent's reason and cannot easily infer it (O'Laughlin & Malle, 2002). A case of pragmatic goals favoring CHR explanations occurs when the explainer downplays the agent's awareness and deliberate decision process and instead points to the "objective" causal determinants of those reasons.

From these two conditions, we can derive an actor-observer asymmetry for reasons versus causal histories. For actors, limited information access will be rare. Actors normally know (or at least believe they know) their reasons for acting and are therefore apt to report them in their explanations (Buss, 1978; Locke & Pennington, 1982). Moreover, since an agent's reasons actually figured in her decision to act, reasons should be highly salient and accessible in the agent's explicit memory (Cowan, 1995; Russell & D'Hollosy, 1992). Observers, by contrast, often do not know the agent's reasons. As a result, they more frequently resort to causal history explanations, citing, for example, a general feature of the action context, a cultural norm, or a personality trait. The information access condition therefore predicts that observers offer a greater number of causal histories (and fewer reasons) than actors do.

Pragmatic goals that favor causal histories can in principle be pursued by either actors or observers. However, more often than not, actors will want to appear as rational decision makers and will therefore prefer reason explanations (Malle et al., 2000) whereas observers don't have a similar impression-management motivation, especially if they have no relationship with the agent. Consequently, pragmatic goals, too, predict that observers offer a greater number of causal histories (and fewer reasons) than do actors.

In addition to delineating multiple modes of explanation, the alternative approach I am proposing also specifies particular features of each mode, and some of

those reflect additional actor-observer asymmetries. For three of the modes—causes, causal history factors, and enabling factors—the features are relatively simple because we are dealing with causal factors that can be classified along traditional attribution dimensions (e.g., internal-external, stable-unstable), allowing a further test of the classic actor-observer hypothesis. More interesting, however, are the specific features of reason explanations, because this mode represents a major departure from classic attribution models.

Reasons have the most complex features of all explanation modes because they are based on the unique conceptual assumptions about how intentional action is generated—namely, by subjective, rational consideration of desires and beliefs and a subsequent decision to act. Two features of reasons stand out (Malle, 1999).

First, any reason appears as one of three mental state types: a desire (“because she wanted more money”), a belief (“because she thought it would help”), or a valuing (“because she found it interesting”). We call this feature *reason type*. Second, explainers either linguistically mark a reason as a subjective mental state by using a mental state verb (such as “I wanted,” “she thought,” “he liked”) or else they leave the reason unmarked. This feature captures the presence or absence of *mental state markers*. One additional feature, the *reason content* (i.e., *what* is desired, believed, or valued), can be classified in the traditional way as being about the agent or the situation. However, this classification appears to have little predictive power (for either actor-observer asymmetries or other phenomena we have examined; Nelson & Malle, 2005; O’Laughlin & Malle, 2002), so I will set it aside in this discussion.

With regard to the two features of reason type and mental state markers, we can derive two further predictions—the belief-desire hypothesis and the unmarked belief hypothesis—regarding actor-observer asymmetries in behavior explanations.

Hypothesis 2: Belief-Desire Asymmetry

This hypothesis posits that actors use more belief reasons and fewer desire reasons than observers do. The principle of information access is critical here as well. When observers try to infer an agent’s reasons, they should have particular difficulties inferring belief reasons, because beliefs often represent idiosyncratic, context-specific information, such as perceived circumstances, anticipated outcomes, and considered alternative paths of acting, which are difficult to know for most observers. Desire reasons, by contrast, can be more easily inferred from general social rules and cultural practices (Bruner, 1990), which is also why children learn to attribute desires before they learn to attribute beliefs (e.g., Nelson-LeGall, 1985; Wellman & Woolley, 1990) and explain actions with desires before they explain them with beliefs (Bartsch & Wellman, 1989). A second psychological process that leads to the belief-desire asymmetry is impression management, which refers to attempts at influencing an audience’s impression of the behavior or agent in

question. Beliefs, but not desires, portray the agent as rational and deliberate (Malle et al., 2000), and actors are normally more concerned than observers with presenting such an image.

Hypothesis 3: Belief Marker Asymmetry

This hypothesis posits that actors, more than observers, leave their belief reasons unmarked (i.e., do not formulate the reason with a mental state verb such as *think*, *believe*, or *know*). Two converging processes suggest this prediction. First, actors' information access to their belief reasons is such that in their minds they directly represent the *content* of their belief—e.g., REPRESENT→[the plants are dry]. They do not normally represent their own belief *qua* mental state; that is, they usually do not REPRESENT→[*I believe* the plants are dry]. When formulating their belief reasons in language, actors will often express simply what they represented, and so they usually leave their belief reasons unmarked: “Why did you turn on the sprinkler?”—“Because the plants were dry.” Observers, by contrast, often represent the agent's thinking *qua* mental state—e.g., REPRESENT→[*she thought* the plants were dry]—and are therefore more likely to formulate a belief reason marked by a mental state verb: “Perhaps *she thought* the plants were dry.”

The second process suggesting a belief marker asymmetry results from the explainer's pragmatic goal of conveying an evaluative attitude toward the agent's reason (Malle, 1999; Malle et al., 2000). Specifically, omitting a belief marker indicates an endorsement of that belief whereas using a marker indicates distance from the belief. For example, if an explainer says, “She turned on the sprinkler because the plants were dry,” he himself seems to believe that the plants were actually dry. By contrast, if he says, “She turned on the sprinkler because *she thought* the plants were dry,” he distances himself from her belief. By explicitly stating that the agent thought the plants were dry, the explainer suggests that there is some doubt as to the truth of the agent's belief. Actors can use this same linguistic device to distance themselves from their own past reasons (“I locked the door because I thought you had already left”), but this is a less frequent case. On the whole, observers are more likely than actors to use belief markers as a distancing device.

I should note that this same logic does not hold for desire reasons because, at least in English, the grammatical form of desires renders marked and unmarked expressions similar to one another (“Why is she running?”—“She wants to be on time” vs. “To be on time”). Because they are so similar, they fail to carry an attitude implication of either distancing or endorsement (Malle et al., 2000).

We now have derived three hypotheses from our alternative theory of explanation:

H1 Actors offer more reasons (and fewer causal history explanations) than observers do.

H2 Actors offer more belief reasons (and fewer desire reasons) than observers do.

H3 Actors offer more unmarked belief reasons than observers do.

What evidence is there in support of these hypotheses?

Evidence for Alternative Predictions

In an extensive series of studies, my colleagues and I have examined actor and observer explanations of behavior in a variety of contexts, including memory protocols, natural interactions, and interviews (Knobe & Malle, 2002; Malle, 2002; Malle, Knobe, & Nelson, 2005). Some studies asked people to recall why-questions and their corresponding explanations, others identified spontaneous explanations in conversation. Some studies let people choose the behaviors they explained, others predetermined those behaviors. In all of them, we let people explain behaviors in their own words and then carefully coded those explanations into conceptual and linguistic categories (Malle, 2003). In total, we examined over 700 participants and over 6,000 explanations.

Hypothesis 1, postulating a reason/CHR asymmetry, received consistent support. Averaged across five studies, actors offered 80% reasons (and 20% CHRs) per intentional behavior explained, whereas observers offered 60% reasons (and 40% CHRs), for a total effect size of $d = 0.69$. Sample reason explanations by actors and sample causal history explanations by observers illustrate this asymmetry:

Reason explanations by actors:

Why did you plan to buy your brother a video game?—'Cause it was his **birthday** [reason] and I **wanted to have him get something that he would actually use** [reason].

Why did you decide to go hiking up the peak?—**Just to kind of get a change of scenery** [reason]. I **was getting sick of what everyone else was doing, kind of** [reason].

Causal history of reason explanations by observers:

Her family like had these family get-togethers all the time 'cause **they had quite a few aunt and uncles in the family** [CHR].

Sometimes he [divorced father] would break in and put gifts all around the house, **because for a long time he was torn between still loving her and hating what she'd done** [CHR].

Hypothesis 2, postulating a belief asymmetry, received consistent support as well. On average, actors offered 65% belief reasons whereas observers offered 45% belief reasons, for an effect size of $d = .52$. Sample belief reasons by actors and sample desire reasons by observers again put this asymmetry into relief.

Belief reasons by actors:

I wanted to pick up a friend 'cause he was just at home doing nothing [belief].

And I looked in this pack and got out this little band-aide case in this pack or something like that, 'cause I thought it would have something that I needed in it [belief].

Desire reasons by observers:

Dad refused to move out for two months after [wife/mother] filed for divorce just because he wanted to be with his kids [desire] and he didn't want to leave the house we had in West Linn [desire].

So she put the knives under her bed to protect herself [desire].

Finally, hypothesis 3 received support as well. On average, actors offered 81% unmarked belief reasons whereas observers offered 60% unmarked belief reasons, $d = 0.58$.

Unmarked belief reasons by actors:

I'm probably taking next fall off because my boyfriend lives in Florida and that in itself is a very difficult situation [unmarked beliefs]

Marked belief reasons by observers:

I'm guessing that he waited [to die] till I left 'cause he knew that I was leaving [marked belief].

All three hypotheses that were derived from our alternative theory of explanation received considerable support across multiple studies and multiple methodologies. In those same studies we were also able to test the two versions of the classic actor-observer asymmetry—one for the general person-situation dichotomy and one for trait explanations. In order to make the two tests independent, we formed two statistical contrasts: the first comparing all person factors to all situation factors, the second comparing traits to nontraits among person factors.

Across these studies, there was no evidence for a person-situation asymmetry among cause explanations of unintentional behavior (actors = 64% person causes, observers = 62% person causes) or among causal history explanations of intentional behavior (actors = 65% person CHRs, observers = 78% person CHRs), with only one out of ten comparisons statistically significant in the predicted direction. Classifying the contents of reason explanations provided no evidence either, as actors mentioned 48% and observers mentioned 50% person content in their reasons.

The trait asymmetry fared only slightly better. Across causal history and cause explanations observers used slightly more traits (38%) than actors did (28%). However, the trait asymmetry held only when the observer knew the agent well (cf. Hampson, 1983).

Why Do Explanations for Self and Other Differ?

We have considered three novel actor-observer hypotheses and found reliable evidence for each of them. Now the question becomes why these asymmetries exist. Our hypotheses generally drew on two psychological principles—information access and pragmatic goals. I now analyze in more detail what these principles entail and hope to clarify what might be special, if anything, about the way the “self” explains behavior.

Hypothesis 1: Reason Asymmetry

If information access is driving the reason asymmetry, there are at least two versions of access we need to consider. The first concerns a direct path of information activation that allows actors to recall the very reasons for which they acted (Brewer, 1994; Malle, 2004). Especially when the actor had engaged in active deliberation, her representation of the action will be intimately tied in memory to her representations of the reason contents in light of which and on the grounds of which she acted.³ For the observer, no such direct connection normally exists (unless the actor shared her deliberations with the observer). Observers rely on a second kind of information access, namely, knowledge structures that contain representations of the circumstances (if the observer was copresent) as well as general knowledge about the actor and the given type of action. These representations can form the basis of reason explanations (e.g., in the case of inferring the actor’s desires and beliefs in the particular circumstances) but will often form the basis of causal history explanations (e.g., in the case of general knowledge about actor or action). The result should be a tendency for observers to use more causal history explanations.

The second process that may be driving the reason asymmetry concerns the explainer’s pragmatic goals, particularly impression management goals, which are often more pronounced for actors than for observers. From previous research we know that reasons help an actor portray herself in a rational light (Malle et al., 2000) whereas CHR explanations keep a “distance” from the agent (Nelson & Malle, 2005), so unless an observer wants to explicitly bridge this distance and make the agent look socially desirable, we would expect actors to offer more reasons than observers do.

A study we conducted recently provide direct evidence for the impression management influence on the reason asymmetry. We found that observers who are explicitly instructed to make the agent look good substantially increased the number of reasons and decreased the number of causal histories in their explanations, with the result that the actor-observer asymmetry was virtually eliminated (Malle et al., 2005). We also have evidence that merely better general knowledge about the agent does not affect the reason asymmetry (Malle et al., 2005). One might suspect that it is primarily the “direct access” difference between actors and observers that influences the asymmetry, but no empirical evidence is currently

available for this hypothesis. One method of testing it is to vary the actor's amount of explicit deliberation before acting or the time passed between action deliberation and explanation, which should decrease the actor's direct memory access to reasons but leave an observer's information basis unchanged. Another method is to make the observer privy to the actor's deliberations, such as by having actors deliberate their action decisions out loud in the laboratory.

Hypothesis 2: Belief-Desire Asymmetry

The two processes of information access and pragmatic goals influence actors and observers differently when it comes to offering belief reasons versus desire reasons. Observers will tend to offer more desire reasons both because desires are easier to infer from the action itself, from its context, and from cultural scripts (Bruner, 1990) and because observers are rarely concerned with presenting the agent in a rational, reasonable light, for which belief reasons are most suitable. If actors, by contrast, know their reasons at all, they know both beliefs and desires, so the choice between belief reasons and desire reasons is solely one of pragmatic goals. Specifically, actors will engage in impression management, which in most situations is better accomplished by beliefs than by desires (Malle et al., 2000). Moreover, they may sometimes engage in "audience design," which is the tailoring of a communicative message—an explanation in this case—to what the audience presumably does not know (e.g., Slugoski, Lalljee, Lamb, & Ginsburg, 1993). Audiences are always in the observer role, and if observers often know (or can infer) desire reasons, then an actor would be more likely to offer belief reasons when explaining her action to an audience.

Recent studies provide initial support for these considerations, though questions linger. In one study we were able to eliminate the established belief-desire asymmetry by comparing actors' explanations to explanations given by observers who knew the agent well (Malle et al., 2005). However, a follow-up study that more closely compared observers' explanations for two groups of agents—strangers and intimates—did not replicate the expected difference in belief reasons. A feature that may account for the inconsistency is that in the study where the actor-observer asymmetry was eliminated, observers were not only very familiar with the agent but were also copresent when the agent performed his or her action. In the second study, by contrast, the intimate observer was not copresent in a good proportion of the cases. We are currently exploring the possible role of copresence as a mediator of observers' access to belief reason information.

The impression management process was tested in one study that specifically instructed observers "to create a positive impression" of the agent. On this instruction, observers increased their use of belief reasons to a rate that was almost as high as actors' rate. However, they also increased their use of desire reasons, which casts doubt on the role of impression management in the *differential* use of belief reasons. Additional studies will have to create a more realistic context for

observers to portray an agent in a positive light, such as making the explainer a spokesperson for the agent and having the audience assess the explainer's successful (vicarious) impression management.

Hypothesis 3: Belief Marker Asymmetry

Actors, I said earlier, tend to represent the contents of their beliefs directly ("It's going to rain.") and may add mental state markers only if they are needed in communication ("Well, I *thought* it would rain."). Observers, by contrast, will normally represent the agent's beliefs *as* mental states ("She thinks it's going to rain.") and therefore readily include a mental state verb. This may be a difficult hypothesis to test, but one way is to have actors explain out loud a number of actions as they are deciding on them, in which case they should report directly the unmarked belief reasons they have on their minds. Observers, by contrast, who watch those same actors decide and are later asked to explain their actions are more likely to provide beliefs with mental state markers.

Probably the more powerful influence on the belief marker asymmetry is the now familiar self-other difference in impression management motivation. An actor will often leave her belief reasons unmarked because that way they sound like statements of fact (e.g., "I watered the azaleas because *the soil was dry!*"), whereas observers normally don't have that kind of impression management motivation. On the contrary, they may sometimes be actively motivated to distance themselves from the agent's beliefs and therefore use a mental state marker to highlight that distance ("She watered the azaleas because *she thought* the soil was dry.>").

Traits Versus Nontraits

The only traditional actor-observer asymmetry for which we have found some evidence in our studies is the greater use of traits by observers than actors. However, the pattern we found lent only partial support to classic claims about trait explanations. First off, there is no evidence for "rampant dispositionism" (L. Ross & Nisbett, 1991) among observers, as they use far fewer trait explanations than nontrait explanations. Expressed as a proportion of all their explanations, observers use traits in less than 10% of cases (and actors in around 5%). Second, the trait asymmetry tended to be more reliable for explanations of unintentional behavior than for explanations of intentional behavior. When accounting for intentional behaviors with causal history explanations, even actors offered a fair number of traits among their person causal history factors ($M = 30\%$), no fewer than observers did ($M = 29\%$). That was especially true in the case of socially undesirable actions, for which both actors and observers normally increase causal history explanations (Nelson and Malle, 2005).

When we turn to the processes that may underlie the trait asymmetry (for unintentional behavior), we must consider Jones and Nisbett's (1972) widely cited postulate of attention differences. However, there is virtually no evidence

that attentional salience leads to more trait explanations (e.g., Bierbrauer, 1976; Enzle & Hansen, 1976; Martin & Huang, 1984; McArthur & Post, 1977; Taylor & Fiske, 1975; Uleman, Miller, Henken, Riley, & Tsemberis, 1981), nor is there any psychological mechanism that would automatically translate attention to an agent's *behavior* into explanations using the agent's *traits* (Knobe & Malle, 2002). Beyond that, Jones and Nisbett suggested that observers use traits because they have less information available than actors do. But when observers have increased knowledge of the agent, they do not use fewer but *more* traits (Hampson, 1983; Kerber & Singleton, 1984; Malle et al., 2005). In fact, observers barely use traits at all unless they know the person well.

To explain the full pattern of data we have available, we must turn Jones and Nisbett's proposal around and posit that reasonable information about the agent's personality or past behavior is a *prerequisite* for observers' trait explanations. Normally observers will, just as actors, try to provide the specific causal trigger of the unintentional behavior in question. When that trigger is unknown, observers who have personality information available can offer traits in place of it. For example:

Why is she in such a good mood today?—**She is always enthusiastic.**

The observer knows the agent and even saw her that day; nonetheless, he doesn't exactly know why she is in such a good mood, and he therefore confines himself to a trait explanation. Actors, by contrast, usually have a pretty good idea what triggered their unintentional behaviors, and they cite those presumed triggers in the form of nontrait cause explanations.

Why are you in such a good mood today?—**Because I had a good interview this morning.**

The considerations must differ in the case of causal history explanations, where actors appear to offer no fewer traits than observers do. Actors, we might suspect, will choose an occasional trait CHR if it helps their impression management. In fact, qualitative inspection of our data suggests that actors may use those traits primarily when they explain actions that lie far in the past and were undesirable, putting their past self down in the service of managing impressions of the current self (e.g., "I was immature and stupid"). This hypothesis can be put to a more direct test by asking actors and observers to explain actions that vary in desirability and in the amount of time passed since performing them.

CONCLUSIONS AND IMPLICATIONS

This chapter has examined two fundamental questions about self-other asymmetries in behavior explanation: What asymmetries are there, and why do they exist? I

surveyed the 30-year evidence for the traditional hypothesis of a person-situation asymmetry and the results were surprising: The meta-analytic average of published studies is very small and the effect is limited to negative events. However, this does not mean that no general self-other asymmetries exist at all; they are just not tracked by the person-situation distinction. An alternative approach to the study of self-other asymmetries relies on what I have called the “folk-conceptual theory of behavior explanation” (Malle, 1999, 2001, 2004). This theory identifies people’s own (i.e., folk-conceptual) assumptions about behaviors and their explanations and postulates two central psychological processes—information access and pragmatic goals—that guide people’s choice of explanations. Three actor-observer hypotheses can be derived from this model: Compared to observers, actors (a) use more reasons and fewer causal histories; (b) more belief reasons and fewer desire reasons; and (c) more unmarked than marked belief reasons. The evidence for these asymmetries is new but consistent across multiple studies (Malle et al., 2005). Evidence for the specific psychological processes that account for each of these asymmetries is as yet incomplete, but the emerging data suggest that all three asymmetries are driven by differences both in information access and impression management. Table 8.3 provides a summary contrast between the traditional and the folk-conceptual view on explanations.

Two questions linger. First, is the traditional actor-observer asymmetry just a myth? Second, what are the implications of the proposed alternative view on self-other asymmetries in explanation?

Is the Traditional Asymmetry a Myth?

Even though the average effect size of the traditional Jones-Nisbett hypothesis amounted to an r of less than .04, it is reasonable to assume that several past studies that found an effect captured real psychological phenomena. One such phenomenon is, of course, the self-serving bias in people’s accounts of negative and positive outcomes, which at times was measured explicitly (and yielded a positive effect size) and at times was confounded with the purportedly tested general actor-observer asymmetry. But my proposal goes further in suggesting that many past studies that found an effect—either the one predicted by Jones and Nisbett or its reversal—may have unwittingly tapped into one or more of the actor-observer asymmetries that the folk-conceptual theory describes and therefore only appeared to demonstrate an actor-observer difference on their person-situation measures.

For example, some studies that found a reverse person-situation difference in explanations of admirable actions (e.g., Mitchell, 1985; Newman, 1978; Schlenker, Hallam, & McCown, 1983) may actually have tapped into the reason asymmetry. Actors who perform an admirable action normally offer reason explanations (Nelson & Malle, 2000), but being faced with “person” and “situation” rating scales they would likely favor the person scale because it better expresses the intentionality and “ownership” they will want to indicate about an admirable action.

TABLE 8.3 Contrast Between Traditional and Folk-Conceptual Views on Actor-Observer Asymmetries in Behavior Explanation

	Traditional Attribution View	Folk-Conceptual View
Events Explained	All events are explained the same way.	Intentional events are explained differently from unintentional events.
Relevant Parameters of Explanation	All explanations are classified into either person/disposition causes or situation causes.	For unintentional behaviors: Causes (person-trait, person-nontrait, situation, interaction). For intentional behaviors: Reasons (belief marked or unmarked; desire marked or unmarked; valuing); Causal Histories (person-trait, person-nontrait, situation, interaction).
Asymmetries	Actors, compared to observers, use more situation causes and fewer disposition causes.	Actors, compared to observers, use (a) more reasons than causal histories, (b) more belief reasons than desire reasons, and (c) more unmarked than marked belief reasons.
Processes Underlying Asymmetries	Information availability, attention focus, sense of freedom.	Information access (direct recall, knowledge structures); pragmatic goals (audience design, impression management)

Other studies coded free-response explanations into person/situation categories and found supportive evidence for the classic thesis because they may have capitalized on the belief asymmetry and belief marker asymmetry identified earlier. How so? Researchers appeared to classify explanations as “about the person” or “about the situation” merely on the basis of the explanation’s *linguistic surface* (Antaki, 1994; Malle, 1999; Miller, Smith, & Uleman, 1981; Monson & Snyder, 1977; L. Ross, 1977). For example, Nisbett, Caputo, Legant, & Marecek’s (1973) illustration of a person attribution for a choice of major was “because I want to make a lot of money” (p. 158) and McGill’s (1989) example was “I like jobs that are challenging” (p. 191). Both of these “person attributions” are marked reasons that mention the agent in the mental state verb. Nisbett et al.’s (1973) illustration of a situation attribution was “because chemistry is a high-paying field” and McGill’s (1989, p. 191) example was “because finance is very challenging.” Both of these “situation attributions” are unmarked beliefs that mention the reason’s situation content. Extrapolating from these examples, we can assume that the category of person attributions picked up on marked reasons (which are often desires; Malle et al., 2000), whereas the category of situation attributions picked up on unmarked reasons with situation content (which are often beliefs; Malle et al., 2000).

Now, we know that when observers offer reasons, they more often cite desire reasons, which, for grammatical reasons, are often marked (Malle, 1999), leading primarily to “person attribution” codes for observer explanations. We also know

that actors use many belief reasons, especially unmarked beliefs (which often have situation content; Malle, 1999), and all that results in a good number of “situation attribution” codes for actor explanations. Thus, if free-response behavior explanations are coded for their linguistic surface, a spurious person-situation asymmetry can emerge, not because of any true self-other asymmetry in person/situation “causes” but because of the confluence of two asymmetries described by the folk-conceptual approach, that of belief reasons and that of belief markers. (For empirical support of these considerations, see Malle, 1999; Malle et al., 2000.)

There may then be some evidence for genuine self-other asymmetries hidden in the bulk of traditional studies (both those that contradicted and those that supported Jones and Nisbett), but it is quite probably evidence for the multitude of folk-conceptual asymmetries I have discussed, not for the broad formulation of a single person-situation difference.

Implications of the Alternative Approach

The alternative view on actor-observer asymmetries stresses the sophisticated conceptual framework that people bring to the task of explaining behavior. Distinctions between types of behavior, modes of explanations, and features of each of mode create a complex set of choices for the explainer. Multiple psychological processes guide those choices, and they can be grouped into two main categories: information access (including direct recall, knowledge structures, representational form) and pragmatic goals (including audience design, impression management, self-distancing). If one acknowledges that this folk-conceptual approach makes more theoretical sense and better fits the data than the classic person-situation model, there is still the question how far the folk-conceptual theory can be pushed. What implications does it have for phenomena such as relationships, perspective taking, or conflict resolution?

Attribution theory has played a notable role in the study of relationships (e.g., Bradbury & Fincham, 1990; Fincham, Bradbury, & Grych, 1990). However, because of the limited predictive power of traditional attribution concepts researchers tended to add more and more “attribution dimensions” to their studies, ranging from locus, stability, and globality to intentionality and responsibility (e.g., Fincham, Beach, & Nelson, 1987). Even so, evidence for the classic actor-observer asymmetries has remained inconsistent (e.g., Fincham, Beach, & Baucom, 1987). A folk-conceptual approach to explanations in relationships would examine naturally occurring behavior explanations that people give for their own and their partner’s behavior and track the various actor-observer asymmetries (for reasons, beliefs, etc.) as indexes of how wide the gap is between self and other in a given relationship. The underlying processes of information access and pragmatic goals could also be independently assessed or manipulated to determine whether self-other gaps are a matter of epistemic, communicative, or motivational problems, and this host of variables could be related to measures of relationship quality and stability.

In romantic as well as political relationships, bridging the self-other gap is a major goal of conflict resolution, and perspective taking is one promising tool with which to achieve this goal. But how do we know when conflict partners do or do not take each other's perspective? Explanations of each other's behaviors could be used as indicators of perspective taking (Nelson, 2003), and changes in actor-observer asymmetries could specifically indicate progress in perspective taking. By monitoring the explanations conflict partners give for their own and the other's behavior before and after conflict-resolution interventions, the effectiveness of such interventions could be assessed. Moreover, the shaping of explanations as having "actor style" or "observer style" may itself serve as a conflict resolution intervention. If actors were encouraged to offer causal history explanations they might gain more distance from their subjective (self-enhancing) point of view, and if observers were encouraged to offer reason explanations they might better see the actor's point of view. Perhaps explaining behavior in the style of the other perspective can help not only take that perspective but also sway the associated evaluations, introducing self-criticalness to the actor perspective and lessening blame from the observer perspective.

The folk-conceptual theory's rich set of parameters also invites studying the developmental dynamics of self-other asymmetries in explanation. With very few exceptions, these asymmetries have not been studied among age groups other than young adults. We might wonder, for example, whether preschool children show any such asymmetries. Some theorists have argued that, at this age, there is no difference in cognition of self and other (Gopnik, 1993), so there should be no cognitively mediated differences in explanations as well. Others maintain that even preschoolers have privileged access to their own mental states (Goldman, 1993), and if so, this privilege should foster both a reason asymmetry and a belief asymmetry in behavior explanations for self and other. Furthermore, we might wonder whether adolescents have renewed difficulties with connecting to other minds, which would amplify certain self-other asymmetries in behavior explanation. Finally, does old age come with the wisdom of greater balance between explanations for one's own and other people's behavior? Questions such as these can be readily studied within the folk-conceptual framework by analyzing naturally occurring explanations and assessing their fundamental parameters at the conceptual, cognitive, and linguistic level.

The search for universals and cultural differences can also be conducted with more acuity by examining parameters of folk explanation at both the linguistic and the conceptual level. Current cross-cultural research on behavior explanations uses the simple person-situation dichotomy for an undoubtedly complex phenomenon (e.g., Choi, Dalal, Kim-Prieto, & Park, 2003; Morris & Peng 1994), and it seems likely that a framework with finer distinctions will be more sensitive to those complexities. For example, claims about a lesser gap between self and other in Eastern cultures can be put to a better test when we distinguish the separate contributions of information access and impression management on a variety of explanation parameters, such as the choice of events explained (intentional vs.

unintentional), the modes of explanation (e.g., reasons vs. causal histories), and the features of each mode (e.g., marked vs. unmarked reasons). Current models depict Eastern explainers as using more “situational attributions,” but the vagueness of this theoretical category allows too many ways in which such findings can come about. For example, the results would be much less impressive if they stemmed from Eastern explainers applying, for pragmatic reasons, fewer mental state markers than if they reflected a cognitive shift toward using more causal history explanations from the self perspective.

In conclusion, the folk-conceptual approach to self-other asymmetries in behavior explanation is not only a tool for critical analysis of past attribution research but also opens numerous avenues for future research. Some of this research will refine the proposed theoretical model of explanation, some will make adjustments to the empirical claims voiced in this chapter. All of it, I hope, will help revise and expand our knowledge of self-other asymmetries in explanation, distancing ourselves from the myths, and bringing us closer to reality.

ACKNOWLEDGMENT

Many thanks for Joshua Knobe and the volume editors for their insightful comments on an earlier draft of this chapter.

NOTES

1. Watson's (1982) review was published long ago, and even at its time the article covered only a limited proportion of the relevant studies.
2. Analyses of other potential moderator variables and details of the various effect size computations can be found in Malle ([submitted](#)).
3. Some readers might ask whether Nisbett and Wilson (1975) have not shown that people normally lack access to the reasons of their actions. But these researchers have shown no such thing (nor has recent research on automaticity). By folk definition, reasons are on the agent's mind at the time of acting, and if the agent does not forget, she can later report on them. Whether there are other, perhaps unconscious processes that helped bring about the action is a separate issue. These additional processes can be cited, of course, in causal history of reason explanations, which are what Nisbett and Wilson offered for their participants' behavior. Their famous studies showed that agents can be unaware of some factors in the causal history of their reasons (hence, their actions), and this corresponds to our position that CHR explanations do not presuppose the agent's subjectivity or awareness.

REFERENCES

- Antaki, C. (1994). *Explaining and arguing: The social organization of accounts*. Thousand Oaks, CA: Sage Publications.
- Aronson, E. (2002). *The social animal* (8th ed.). New York: Worth Publishers.
- Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*, 708–717.
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.
- Bierbrauer, G. (1979). Why did he do it? Attribution of obedience and the phenomenon of dispositional bias. *European Journal of Social Psychology, 9*, 67–84.

- Bradbury, T. M., & Fincham, F. D. (1990). Attributions in marriage: Review and critique. *Psychological Bulletin*, *107*, 3–33.
- Bretherton, I. (1991). Intentional communication and the development of an understanding of mind. In D. Frye & C. Moore (Eds.), *Children's theories of mind* (pp. 49–75). Hillsdale, NJ: Erlbaum.
- Brewer, W. F. (1994). Autobiographical memory and survey research. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 11–20). New York: Springer.
- Bruner, J. (1990). Acts of meaning. Cambridge, MA: Harvard University Press.
- Buss, A. R. (1978). Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology*, *36*, 1311–1321.
- Choi, I., Dalal, R., Kim-Prieto, C., & Park, H. (2003). Culture and judgment of causal relevance. *Journal of Personality and Social Psychology*, *84*, 46–59.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, *60*, 685–700.
- Enzle, M. E., & Hansen, R. D. (1976). Effects of video-mediated visual contact on observers' attributions of causality and reciprocal game behavior. *Simulation and Games*, *7*, 281–294.
- Fincham, F. D., Beach, S. R., & Baucom, D. H. (1987). Attribution processes in distressed and nondistressed couples: 4. Self-partner attribution differences. *Journal of Personality and Social Psychology*, *52*, 739–748.
- Fincham, F. D., Beach, S. R., & Nelson, G. (1987). Attribution processes in distressed and nondistressed couples: III. Causal and responsibility attributions for spouse behavior. *Cognitive Therapy and Research*, *11*, 71–86.
- Fincham, F. D., Bradbury, T. N., & Grych, J. H. (1990). Conflict in close relationships: The role of intrapersonal phenomena. In S. Graham & V. S. Folkes (Eds.), *Attribution theory: Applications to achievement, mental health, and interpersonal conflict* (pp. 161–184). Hillsdale, NJ: Erlbaum.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (second ed.). New York: McGraw-Hill.
- Goldman, A. I. (1993). The psychology of folk psychology. *Behavioral and Brain Sciences*, *16*, 15–28.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*, 1–14.
- Greenberg, M. S., & Frisch, D. M. (1972). Effect of intentionality on willingness to reciprocate a favor. *Journal of Experimental Social Psychology*, *8*, 99–111.
- Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, *35*, 603–618.
- Hampson, S. E. (1983). Trait ascription and depth of acquaintance: The preference for traits in personality descriptions and its relation to target familiarity. *Journal of Research in Personality*, *17*, 398–411.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Jones, E. E. (1976). How do people perceive the causes of behavior? *American Scientist*, *64*, 300–305.
- 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* (pp. 79–94). Morristown, NJ: General Learning Press.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 129–238). Lincoln: University of Nebraska Press.
- Kerber, K. W., & Singleton, R. (1984). Trait and situational attributions in a naturalistic setting: Familiarity, liking, and attribution validity. *Journal of Personality*, *52*, 205–219.
- Knobe, J., & Malle, B. F. (2002). Self and other in the explanation of behavior: 30 years later. Special issue on self-other asymmetries. *Psychologica Belgica*, *42*, 113–130.
- Krueger, J. (1998). On the perception of social consensus. *Advances in Experimental Social Psychology*, *30*, 163–240.
- Kugelmass, S., & Breznitz, S. (1968). Intentionality in moral judgment: Adolescent development. *Child Development*, *39*, 249–256.
- Locke, D., & Pennington, D. (1982). Reasons and other causes: Their role in attribution processes. *Journal of Personality and Social Psychology*, *42*, 212–223.

- Locke, K. D. (2002). Are descriptions of the self more complex than descriptions of others? *Personality and Social Psychology Bulletin*, 28, 1094–1105.
- Malle (2005). The actor-observer asymmetry in causal attribution: A (surprising) meta-analysis. Manuscript submitted for publication.
- Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, 3, 23–48.
- Malle, B. F. (2001). Folk explanations of intentional action. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 265–286). Cambridge, MA: MIT Press.
- Malle, B. F. (2002). The social self and the social other. Actor-observer asymmetries in making sense of behavior. In J. P. Forgas & K. D. Williams (Eds.), *The social self: Cognitive, interpersonal, and intergroup perspectives* (pp. 189–204). Philadelphia, PA: Psychology Press.
- Malle, B. F. (2003). *FEx: Coding scheme for people's folk explanations of behavior*. University of Oregon. Retrieved from <http://darkwing.uoregon.edu/~bfmalle/fex.html> on July 1, 2004.
- Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.
- 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., & Nelson, S. E. (2003). Judging mens rea: The tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences and the Law*, 21, 1–18.
- Malle, B. F., & Pearce, G. E. (2001). Attention to behavioral events during social interaction: Two actor-observer gaps and three attempts to close them. *Journal of Personality and Social Psychology*, 81, 278–294.
- Malle, B. F., Knobe, J., & Nelson, S. (2005). *Actor-observer asymmetries in folk explanations of behavior: New answers to an old question*. Manuscript in preparation, University of Oregon.
- Malle, B. F., Knobe, J., O'Laughlin, M., Pearce, G. E., & Nelson, S. E. (2000). Conceptual structure and social functions of behavior explanations: Beyond person–situation attributions. *Journal of Personality and Social Psychology*, 79, 309–326.
- Martin, D. S., & Huang, M. (1984). Effects of time and perceptual orientation on actors' and observers' attributions. *Perceptual and Motor Skills*, 58, 23–30.
- McArthur, L. Z., & Post, D. L. (1977). Figural emphasis and person perception. *Journal of Experimental Social Psychology*, 13, 520–535.
- McClure, J., & Hilton, D. (1997). For you can't always get what you want: When preconditions are better explanations than goals. *British Journal of Social Psychology*, 36, 223–240.
- McGill, A. L. (1989). Context effects in judgments of causation. *Journal of Personality and Social Psychology*, 57, 189–200.
- Mele, A. R. (1992). *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.
- Miller, F. D., Smith, E. R., & Uleman, J. (1981). Measurement and interpretation of situational and dispositional attributions. *Journal of Experimental Social Psychology*, 17, 80–95.
- Mitchell, T. E. (1985). Actor-observer differences in attributions to morality. *Journal of Social Psychology*, 125, 475–477.
- Monson, T. C., & Snyder, M. (1976). Actors, observers, and the attribution process: Toward a reconceptualization. *Journal of Experimental Social Psychology*, 13, 89–111.
- Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, 67, 949–971.
- Nelson, S. E. (2003). *Setting the story straight: A study of discrepant accounts of conflict and their convergence*. Unpublished Doctoral Dissertation, University of Oregon
- Nelson, S. E., & Malle, B. F. (2000, April). *Explaining intentional actions: Explanations as modifiers of social perception and judgment*. Poster presented at the Annual Meeting of the Western Psychological Association, Portland, Oregon.
- Nelson, S., & Malle, B. F. (2005). Self-serving biases in explanations of intentional behavior. Manuscript in preparation, University of Oregon.

- Nelson-LeGall, S. A. (1985). Motive-outcome matching and outcome foreseeability: Effects on attribution of intentionality and moral judgments. *Developmental Psychology, 21*, 323–337.
- Newman, A. (1978). Actor-observer differences in perception of self-control. *Journal of Social Psychology, 105*, 199–204.
- Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737–759.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.
- Nisbett, R. E., Caputo, C., Legant, P., & Marecek, J. (1973). Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology, 27*, 154–164.
- O'Laughlin, M. J., & Malle, B. F. (2002). How people explain actions performed by groups and individuals. *Journal of Personality and Social Psychology, 82*, 33–48.
- Read, S. J. (1987). Constructing causal scenarios: A knowledge structure approach to causal reasoning. *Journal of Personality and Social Psychology, 52*, 288–302.
- Robins, R. W., Spranca, M. D., & Mendelsohn, G. A. (1996). The actor-observer effect revisited: Effects of individual differences and repeated social interactions on actor and observer attributions. *Journal of Personality and Social Psychology, 71*, 375–389.
- Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*, 677–688.
- 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*. New York: McGraw-Hill.
- Ross, L., Greene, D., & House, P. (1977). The “false-consensus effect:” An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279–301.
- Ross, M., & Fletcher, G. J. O. (1985). Attribution and social perception. In G. Lindzey & E. Aronson (Eds.), *The Handbook of Social Psychology* (Vol. 2, pp. 73–114). New York: Random House.
- Russell, E. W., & D'Hollosy, M. E. (1992). Memory and attention. *Journal of Clinical Psychology, 48*, 530–538.
- Schlenker, B. R., Hallam, J. R., & McCown, N. E. (1983). Motives and social evaluation: Actor-observer differences in the delineation of motives for a beneficial act. *Journal of Experimental Social Psychology, 19*, 254–273.
- Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Sheldon, K. M., & Johnson, J. T. (1993). Forms of social awareness: Their frequency and correlates. *Personality and Social Psychology Bulletin, 19*, 320–30.
- Slugoski, B. R., Lalljee, M., Lamb, R., & Ginsburg, G. P. (1993). Attribution in conversational context: Effect of mutual knowledge on explanation-giving. *European Journal of Social Psychology, 23*, 219–238.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*, 193–210.
- Taylor, S. E., & Fiske, S. T. (1975). Point-of-view and perceptions of causality. *Journal of Personality and Social Psychology, 32*, 439–445.
- Turnbull, W. (1986). Everyday explanation: The pragmatics of puzzle resolution. *Journal for the Theory of Social Behavior, 16*, 141–160.
- Uleman, J. S., Miller, F. D., Henken, V., Riley, E., & Tsemberis, S. (1981). Visual perspective or social perspective?: Two failures to replicate Storms' rehearsal, and support for Monson and Snyder on actor-observer divergence. *Replications in Social Psychology, 1*, 54–58.
- Watson, D. (1982). The actor and the observer: How are their perceptions of causality divergent? *Psychological Bulletin, 92*, 682–700.
- Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35*, 245–275.