Actor–Observer Asymmetries in Explanations of Behavior: New Answers to an Old Question

Bertram F. Malle
University of Oregon

Joshua M. Knobe
University of North Carolina at Chapel Hill

Sarah E. Nelson
Harvard Medical School

Traditional attribution theory conceptualizes explanations of behavior as referring to either dispositional or situational causes. An alternative approach, the folk-conceptual theory of behavior explanation, distinguishes multiple discrete modes of explanation and specific features within each mode. Because attribution theory and the folk-conceptual theory carve up behavior explanations in distinct ways, they offer very different predictions about actor–observer asymmetries. Six studies, varying in contexts and methodologies, pit the 2 sets of predictions against each other. There was no evidence for the traditional actor–observer hypothesis, but systematic support was found for the actor–observer asymmetries hypothesized by the folk-conceptual theory. The studies also provide initial evidence for the processes that drive each of the asymmetries: impression management goals, general knowledge, and copresence.

Keywords: attribution, self-perception, social perception, social cognition

Supplemental materials: http://dx.doi.org/10.1037/0022-3514.93.4.491.supp

In person perception, people assume two basic perspectives: As *observers* they try to make sense of other people's behavior; as *actors* they try to make sense of their own behavior. This fundamental duality, codified in language as the first-person and third-person forms, is particularly striking in explanations of behavior, where two people may account for the same event in dramatically different ways.

Within traditional attribution theory, explanations of behavior have been conceptualized as referring to either situation causes or dispositional/person causes (Jones & Davis, 1965; Kelley, 1967; for reviews, see Malle, 2004, chap. 1; Ross & Fletcher, 1985; Shaver, 1975). Using this conceptualization, Jones and Nisbett (1972) proposed the classic actor–observer asymmetry in explanation, claiming that "there is a pervasive tendency for actors to attribute their actions to situational requirements, whereas observers tend to attribute the same actions to stable personal dispositions" (p. 80).

This asymmetry is widely accepted in social psychology (e.g., Aronson, 2002; Baron, Byrne, & Branscombe, 2006; Fiske, 2004; Kenrick, Neuberg, & Cialdini, 2006); it is well represented in psychology as a whole (e.g., Davis & Palladino, 2004; Gray, 2002; Lahey, 2003; Meyers, 2004; Rathus, 2004); and actor–observer considerations have reached into other disciplines as well, such as management studies, artificial intelligence, semiotics, anthropology, and political science (Galibert, 2004; Jin & Bell, 2003; Larsson, Västfjäll, & Kleiner, 2001; Marsen, 2004; Raviv, Silberstein, Raviv, & Avi, 2002; Rogoff, Lee, & Suh, 2004).

Not surprisingly, then, the actor–observer asymmetry has been described as "robust and quite general" (Jones, 1976, p. 304), "pervasive" (Aronson, 2002, p. 168), and "an entrenched part of scientific psychology" (Robins, Spranca, & Mendelsohn, 1996, p. 376). It is considered "firmly established" (Watson, 1982, p. 698), as "evidence for the actor–observer effect is plentiful" (Fiske & Taylor, 1991, p. 73).

Unfortunately, the empirical evidence does not support these assertions. A recent meta-analysis of 173 studies in 113 articles on the classic actor–observer asymmetry yielded average effect sizes (d) of −0.015 to 0.095, depending on statistical models and spe-

Bertram F. Malle, Department of Psychology, University of Oregon; Joshua M. Knobe, Department of Philosophy, University of North Carolina at Chapel Hill; Sarah E. Nelson, Division on Addictions, Cambridge Health Alliance, Harvard Medical School.

The research reported in this article was partially supported by National Science Foundation CAREER Award SBR-9703315. We are grateful to the researchers who contributed to the studies reported in this article: Kristen MacConnell, Sam Stevens, Gale Pearce, Matt O'Laughlin, and Katie MacCionnaith. Thanks also go to John Barresi, Lara London, and Dan Rothschild for valuable discussions.

Correspondence concerning this article should be addressed to Bertram F. Malle, Department of Psychology, 1227 University of Oregon, Eugene, OR 97403-1227. E-mail: bfmalle@uoregon.edu

Journal of Personality and Social Psychology, 2007, Vol. 93, No. 4, 491–514
 0022-3514/07/$12.00 DOI: 10.1037/0022-3514.93.4.491

cific attribution scores (Malle, 2006). Corrections for possible publication bias turned the average effect size to 0. Remarkably, whereas a handful of studies that reported evidence in favor of the hypothesis has been cited for decades (e.g., Nisbett, Caputo, Legant, & Marecek, 1973; Storms, 1973), a large number of studies that found no or even opposite evidence never entered the scientific discourse, the qualitative reviews, or the textbooks.

The meta-analysis did reveal a few variables that occasionally push the actor–observer asymmetry away from the zero point. For negative events, for example, the predicted asymmetry held, while for positive events, the opposite held, turning the average valence-corrected effect to zero. Other moderators suggested artifacts in the literature (e.g., the effect held when the explained event was hypothetical or when the actor was portrayed as highly unusual to observers). In addition, 20 studies that assessed explanations not as person–situation rating scales but as content-coded free responses showed an average effect size of 0.32, and this value jumped to 0.42 in those 10 studies that examined only intimates' free-response explanations. Clearly, a new examination of the actor–observer asymmetry must rely on free-response explanations and should clarify whether intimates really do show the effect more so than strangers, contrary to what Jones and Nisbett (1972) had predicted.

But there is a broader concern. The meta-analytic results cast doubt not only on the methodology of standard attribution studies but on the classic hypothesis itself and its underlying theoretical framework. The present studies looked carefully for evidence of the classic asymmetry under the most favorable methodological conditions, using free-response behavior explanations across a variety of contexts, methods, and levels of actor–observer familiarity. A real possibility existed, however, that the predicted findings would not emerge because the very theory of person–situation attributions is incorrect. In parallel to reexamining the classic asymmetry, we therefore tested the predictions of an alternative theory of behavior explanations that makes quite different assumptions about how attributions work (Malle, 1999, 2004, 2006; Malle, Knobe, O'Laughlin, Pearce, & Nelson, 2000). This alternative theory predicts not just one but three actor–observer asymmetries (Knobe & Malle, 2002; Malle, 1999). If these predictions turn out to be correct and the traditional predictions incorrect, then we may conclude that actor–observer asymmetries do indeed exist, but also that they can be uncovered only when we accept a shift to a new theoretical framework.

Across six studies reported in this article we ask these questions: Are there any actor–observer differences in people's behavior explanations? Which theoretical model predicts them? And what psychological processes—such as intimacy or impression management—might account for the differences?

We proceed as follows. First we introduce an alternative model of attribution, the folk-conceptual theory of explanation. Then we generate predictions from this model that directly compete with the traditional actor–observer hypothesis. Next we report six studies that tested both the traditional hypothesis and the three new hypotheses, followed by a meta-analytic summary of the results. Studies 4–6 also identify some of the psychological processes that underlie the asymmetries, and additional predictions are derived for future research.

The Folk-Conceptual Theory of Behavior Explanations

Humans perceive and conceptualize intentional action as a unique natural phenomenon and treat it differently from unintentional behavior, and indeed from any other physical event (Heider, 1958; Malle & Knobe, 1997a). Intentionality detection is grounded in older primate capacities (e.g., Call & Tomasello, 2005; Premack & Woodruff, 1978), emerges early in infancy (e.g., Woodward, 1999), and is a hallmark of the child's developing theory of mind (Gopnik & Meltzoff, 1997; Perner, 1991; Wellman, 1990). Obviously, intentionality lies at the heart of the adult's folk conception of mind and behavior as well (Kashima, McIntyre, & Clifford, 1998; Malle & Knobe, 1997a; Malle, Moses, & Baldwin, 2001), and it therefore provides the fundamental concept of a new theory of behavior explanations that we have developed over the last decade (Malle, 1999, 2004, 2007; Malle et al., 2000; O'Laughlin & Malle, 2002).

The theory posits that people use qualitatively distinct modes of explanation depending on whether they perceive a behavior to be intentional or unintentional (Malle, 1999; see also Abraham, 1988; Buss, 1978; Harré, 1988; Heider, 1958; Lalljee & Abelson, 1983; Locke & Pennington, 1982; McClure, 1984; Read, 1987; White, 1991). These different explanation modes are not merely assorted causes that differ in some attribute (such as locus or stability); rather, they are distinct approaches people take to the problem of explaining behavior, grounded in conceptual assumptions about the nature of the behavior explained (Malle, 2001; Malle et al., 2000). Because these assumptions, especially the intentionality concept itself, are quite complex, the theory posits several distinct modes of explanations. Most apparent are the modes of cause explanations and reason explanations: Events perceived to be unintentional are explained by causes that mechanically brought about the event; those perceived to be intentional are typically explained by the agent's reasons for acting (Audi, 1993; Davidson, 1963; Donellan, 1967; Heider, 1958; Malle, 1999; Mele, 1992; Searle, 1983). Consider the following two explanations.

> (1) "Anne studied for the test all night because *she wanted to do well*."
>
> (2) "Anne was nervous about the test results because *she wanted to do well*."

The explanation clauses in the two sentences, though identical on the linguistic surface, reveal very different assumptions that the explainer makes about the relation between the agent and the behavior. In Example 1, Anne studied *in order* to do well, she *chose* to study, she studied *for the reason* stated, namely, wanting to do well. These assumptions characterize Anne as a thinking, choosing, reasoning agent and the behavior as performed because of Anne's reasoning and choosing to so act. None of these inferences hold in Example 2. There, Anne's nervous state is simply caused by her desire to do well; no reasoning, no planning, no choice is involved, and she may not even be aware of the causal relation. Clearly, reason explanations work very differently from other explanations, and so we need to take a close look at the nature of reasons.

Reason Explanations

Reasons can be defined as the contents of an agent's mental states (primarily beliefs and desires) in light of which and on the

grounds of which the agent formed an intention to act. When people cite a reason explanation for an action, they ascribe to the agent one or more beliefs or desires that (they presume) figured in the agent's decision to so act. This is most obvious when the agent herself provides a reason explanation:

> (3) "Why did you go running?"—"Um, because *I wanted to get in better shape*, and . . . *I figured that I can do that by going running*."

Observers, too, can emphasize (what they presume to be) the agent's own reasons to act:

> (4) "Why did they sell their car?"—"*They felt it was too small for the family*."

Providing a reason is an act of perspective taking because the explainer tries to cite what the agent had on his or her mind when deciding to act. Malle et al. (2000) showed that a reason explanation becomes meaningless if the agent's awareness of the reason is denied, as in "Anne invited Ben for dinner because he had helped her paint her room (even though she was not aware that he had helped her paint her room)." Thus, when people offer reason explanations, they make an *assumption of subjectivity*—they assume that the agent was aware of her reasons and acted on those subjectively held reasons (whether or not they reflect objective facts).

People also make a second assumption, which we can label the *assumption of rationality*. Reasons connect with actions not only by causal force but also by a compelling logic of rationality: Given the agent's beliefs and desires, the intention or action at issue follows by the rules of practical reasoning. Consider Example 3 more schematically:

> X wanted O [to get in better shape].
>
> X believed that A [running] leads to O.
>
> Therefore, X intended to do A.

Philosophers since Aristotle have analyzed the unique nature of practical reasoning, because following this logic is considered a hallmark of rationality. What is important for our purposes is that people explain intentional action in accordance with this practical logic, which reflects their assumption of rationality with respect to reason explanations.

Causal History of Reason Explanations

Even though people explain most intentional behaviors by reference to the agent's reasons, they explain some of them by pointing to factors that lie in the background of those reasons. These factors can be subsumed under the label *causal history of reasons* and include such forces as the agent's unconscious, personality, upbringing, and culture, along with the immediate context (Malle, 1994, 1999; Malle et al., 2000; O'Laughlin & Malle, 2002; see also Hirschberg, 1978; Locke & Pennington, 1982).[1] Whereas reason explanations try to capture what the agent herself considered and weighed when deciding to act, causal history explanations take a step back and try to capture processes that led up to the agent's reasons. For example, when clarifying why Kim didn't vote, an explainer might say, "She is lazy" or "Her whole family is apolitical." These statements provide explanations of an intentional action, but they do not pick out Kim's subjective reasons for not voting. Causal history of reason explanations help explain an intentional action by citing causal antecedents to the agent's reasoning and decision to act, but there is no assumption that the agent actively considered those antecedents in her reasoning process. Hence, when an explainer states, "Kim didn't vote because she is lazy," he does not imply that Kim thought, "I am lazy; therefore I shouldn't vote." Rather, the explainer indicates that Kim's laziness was part of the causal background that gave rise to her reasoning. In short, finding causal history explanations is not an act of perspective taking.

In addition, causal history explanations do not assume any rational connection between the causal history factors and the act to be explained. Laziness, childhood experiences, culture, and other background factors causally contribute to the action (by bringing about relevant reasons), but the background's contribution often lacks rationality, such as in the case of outdated cultural conventions, mindless personal habits, "primes" in the immediate context, or unconscious motives.

Table 1 shows three intentional behaviors, each explained by reasons and causal histories (explanations that people reliably distinguished in a previous study; Malle, 1999). We see that the distinction between reasons and causal history explanations has nothing to do with the classic person–situation distinction. Some causal history explanations refer to the person (e.g., "He is driven to achieve"), whereas others refer to the situation (e.g., "That's the cultural norm"). Likewise, some reasons mention the person ("He thought it would be cool"), whereas others mention the situation ("A project was due"). The features that distinguish reasons from causal history explanations are the assumptions of subjectivity and rationality, which are necessarily present in reason explanations but not in causal history explanations.

According to the folk-conceptual theory of explanations, then, one important choice that people face when explaining intentional behavior is whether to offer reasons or causal histories (or both). However, this is not the only choice.

Types and Features of Reasons

Once people offer a reason explanation, two further choices arise (Malle, 1999): what type of reason to provide (typically either a belief or a desire) and whether to mark this reason with a mental state verb (such as "He wanted" or "She thought").

Beliefs and desires. In people's folk concept of intentionality, both beliefs and desires serve as necessary conditions of an intention to act (Malle & Knobe, 1997a), and both are frequently cited in explanations of intentional action. For example, when explaining why Ian has been working so much lately, one might cite a desire such as "He wants that promotion" or a belief such as "He realizes the project is due in a week." Is there any psychological

[1] Causal history of reason explanations account for intentional behavior and are therefore distinct from cause explanations, which account for unintentional behavior (discussed later in this section). What the two have in common is the mechanism of a simple cause–effect relation, but causal history explanations describe what brought about reasons and therefore intentional behavior, whereas cause explanations describe what brought about unintentional behavior.

Table 1
Reason Explanations and Causal History of Reason (CHR) Explanations for Three Behaviors

Behavior	Reason explanation	CHR explanation
Kim chose not to vote in the last election.	She thought that none of the candidates was trustworthy. [*marked belief reason*]	She doesn't realize that every vote counts. [*person-mental state CHR*]
	She didn't want to support the system. [*marked desire reason*]	She is lazy. [*person-trait CHR*]
By choice, Ian worked 14 hours a day last month.	To make more money. [*unmarked desire reason*]	That's the cultural norm there. [*situation CHR*]
	A project was due. [*unmarked belief reason*]	He is driven to achieve. [*person-trait CHR*]
Brian used heavy drugs last Sunday at the party.	He was curious what it would feel like. [*marked desire reason*]	A bunch of others used them. [*situation-other person CHR*]
	He thought it would be cool. [*marked belief reason*]	He grew up in a drug-dealing home. [*person-situation interaction CHR*]

Note. Adapted from "How People Explain Behavior: A New Theoretical Framework," by B. F. Malle, 1999, *Personality and Social Psychology Review, 3,* p. 35. Copyright 1999 by Erlbaum. The terms *marked* and *unmarked* refer to reason explanations that are expressed either with a mental state marker ("he thought," "she wanted") or not.

significance to the explainer's choice between offering belief reasons and desire reasons? At times it may not matter because one implies the other ("He thinks hard work will get him the promotion" ⇒ "He wants that promotion"). But at other times it matters quite a bit. For one thing, belief reasons, more than desire reasons, provide idiosyncratic details about the agent's decision-making process, including rejected options, specific plans of action initiation, and considered long-term consequences. For another, belief reasons refer to the agent's thinking and knowledge, drawing attention to the agent's rational, deliberative side, whereas desire reasons highlight what the agent wants, needs, and hence lacks (Malle et al., 2000).

Mental state markers. A reason explanation can be linguistically expressed in two different ways. The explainer may use a mental state verb to mark the type of reason cited (i.e., a belief or desire), or the explainer may omit such a verb and directly report the content of that reason. Suppose our explainer is faced with the question, "Why did she go to the Italian café?" If he chose to cite a desire reason, he could use the marked form:

(5) "She went to the café *because she wants to have an authentic cappuccino*."

Or he could use the unmarked form:

(6) "She went to the café [____] *to have an authentic cappuccino*."

Likewise, if the explainer chose to cite a belief reason, he could use the marked belief reason:

(7) "She went to the café *because she thinks they have the best cappuccino*."

Or he could use the unmarked belief reason:

(8) "She went to the café *because* [____] *they have the best cappuccino*."

Marked and unmarked reasons do not express two different hypotheses about why the action was performed; rather, they express the same hypothesis in two different ways. This difference is not trivial, however. Citing or omitting mental state markers can serve significant social functions, both for self-presentation and for conveying one's attitude toward the agent (Malle, 1999; Malle et al., 2000), a topic to which we return shortly. Table 1 provides a number of additional examples of marked and unmarked belief and desire reasons.

Explanations of Unintentional Behavior

So far we have presented the folk-conceptual theory as it pertains to explanations of behaviors that people perceive as *intentional*. For these behaviors, the model postulates a conceptual framework that departs significantly from the framework postulated by traditional attribution theory. By contrast, for explanations of behaviors that people perceive as *unintentional*, the folk-conceptual theory does not fundamentally differ from attribution theory. According to the folk-conceptual model, there is only one mode that explains unintentional behavior—cause explanations—and this mode operates much the same way as explanations of any physical event. Cause explanations of unintentional behavior do not involve any complex conceptual assumptions about intentionality, subjective reasons, or rationality. They simply cite factors that, according to the explainer, brought about the event in question. If needed, these causal factors can be classified along dimensions such as internal–external, stable–unstable, and so on (Peterson, Schulman, Castellon, & Seligman, 1992), and in the domain of outcome attributions these dimensions have proven to be predictively useful (Weiner, 1986).

These similarities in theorizing about explanations of unintentional behavior notwithstanding, the folk-conceptual theory of explanation and classic attribution theory differ substantially in their predictions of actor–observer asymmetries in behavior explanation.

Predictions of Actor–Observer Asymmetries

According to the folk-conceptual theory of explanation, people's explanations of intentional behavior vary meaningfully in three major parameters of explanation: (a) the use of reason explanations versus causal history explanations, (b) the use of

belief reasons versus desire reasons, and (c) the use of mental state markers when referring to reasons. By contrast, in traditional attribution theory, people's explanations are classified into the categories of person (or trait) versus situation attributions. Because the folk-conceptual model and attribution theory carve up explanations in distinct ways, they offer very different tools for predicting actor–observer asymmetries. For illustration, consider the two explanations "[I/She] yelled at him because [___] he broke the window" and "[I/She] yelled at him because it was so hot outside." Attribution theory places these two explanations in the same category, namely, *situation attributions* ("he broke . . ." and "it was hot . . ."). It must therefore predict that actors will be more likely to use both sorts of explanations. By contrast, the folk-conceptual model classifies the first explanation as an unmarked belief reason (when yelling at him the agent is aware that he broke the window, so the fully marked explanation would be ". . . because [I/she] *thought* he broke the window"); but the model classifies the second as a causal history explanation, and it therefore predicts that actors will be more likely to use the first whereas observers will be more likely to use the second type of explanation. We now examine these distinct predictions in detail.

Traditional Attribution Predictions

Traditional attribution theory predicts that actors will make more situational attributions whereas observers will make more dispositional attributions (Jones & Nisbett, 1972). This prediction is equivocal in two respects. First, the term *disposition* is ambiguous (Ross & Fletcher, 1985), as it has been used to refer either to any factor that lies within the person (including emotions, traits, and beliefs) or solely to stable personality traits. Because traits are only one type of person factor, the traditional thesis actually breaks down into two independent contrasts: Actors and observers may differ in their use of (a) person factors versus situation factors and (b) traitlike person factors versus nontraitlike person factors. We test both of these contrasts in our studies.

Second, there is an ambiguity in the domain of behaviors to which the traditional thesis applies. Some researchers have claimed that the causal attribution framework applies to all behaviors, whether intentional or unintentional (e.g., Kelley, 1967; Nisbett et al., 1973); others have claimed it applies to intentional behaviors only (e.g., Jones & Davis, 1965; Shaver, 1975); and still others have claimed that the classic framework applies only to unintentional behaviors (Kruglanski, 1975; Malle, 1999). Our studies test the validity of each of these claims.

The attribution literature has not converged on an account of the psychological processes that are presumed to underlie the traditional actor–observer asymmetry (Knobe & Malle, 2002; Monson & Snyder, 1976; Robins et al., 1996). Jones and Nisbett (1972) posited two main processes to explain person–situation effects: differences in attention and differences in knowledge. But later investigations have called both of these accounts into question. Despite early evidence for an attention account of the actor–observer asymmetry (Storms, 1973), later studies did not replicate this evidence (e.g., Uleman, Miller, Henken, Riley, & Tsemberis, 1981; see Malle, 2006). The knowledge account, too, faced numerous disconfirming studies (e.g., Kerber & Singleton, 1984; Taylor & Koivumaki, 1976), and a meta-analysis unexpectedly shows a stronger asymmetry for familiar observers than for unfamiliar observers (Malle, 2006). The results of the present studies will shed light on the possible accounts of the traditional actor–observer asymmetry, should this asymmetry emerge.

Folk-Conceptual Predictions

Within the folk-conceptual model we can develop three actor–observer asymmetries, one for each of the three major parameters of explanation. As with many phenomena of social cognition, two broad psychological processes combine to bring about these asymmetries: information access and motivation (Barresi, 2000; Knobe & Malle, 2002; Malle, 2004, 2005).

The first prediction is a *reason asymmetry*, which posits that actors use more reasons and fewer causal history explanations (relative to base rates) than observers do. Cognitive access may contribute to this asymmetry, because actors normally know their reasons for acting and are therefore apt to report them in their explanations (Buss, 1978; Locke & Pennington, 1982). Moreover, because the actor's reasons actually figured in the decision to act, reasons should be highly accessible in the actor's memory. Observers, by contrast, normally have no access to the decision process that leads up to the action and must rely on mental simulation, context-specific inference, and general knowledge to construct an explanation, which will more often refer to the causal history of the actor's reasons.

Another psychological process that may contribute to the reason asymmetry is the motivational process of impression management (by which we mean attempts to influence an audience's impression of either oneself or another person). Reason explanations tend to portray the actor as a conscious, rational agent with the capacity to choose (Knobe & Malle, 2002), whereas causal history explanations tend to highlight the causal nexus that impinges on the actor, the forces that are out of the person's control and awareness (Malle et al., 2000). There are many contexts and roles that may influence the direction of impression management, but as a rule, we can expect actors to provide explanations that paint a self-flattering picture (at least within Western cultures; Sedikides & Strube, 1997). Observers, by contrast, will less often make an effort to portray the actor in an especially positive light. Thus, both information access and impression management predict that actors give relatively more reason explanations than observers do.

The second prediction of the folk-conceptual theory is a *belief asymmetry*, which posits that actors use relatively more belief reasons and fewer desire reasons than observers do. Cognitive access should be involved here as well, but not as a function of the actor's privileged access to her reasons (because there should be no tendency to directly recall one type of reason any better than another) but because of a specific limitation on the observer's side. Observers who try to infer an actor's reasons have a particular difficulty inferring belief reasons, because beliefs often represent idiosyncratic perceptions of circumstances, options, and outcomes. If the observer has no knowledge of these idiosyncrasies, it will be easier to infer desire reasons, because they more easily derive from general social rules and cultural practices (Bruner, 1990), are more immediately visible in human movement (Baird & Baldwin, 2001; Phillips & Wellman, 2005), and are more quickly recovered in perceptions of behavior (Holbrook, 2006).

Previous findings suggest that belief reasons can portray the actor as rational and may thus serve impression management

functions (Malle et al., 2000). At the same time, desire reasons can downplay overt deliberateness and support modesty concerns (e.g., "I just wanted to say hi"; "I just wanted to play my best"). The influence of impression management on the belief asymmetry may therefore be weaker and more context sensitive.

The third hypothesis concerns a belief marker asymmetry (henceforth labeled the *marker asymmetry*), which posits that actors leave their belief reasons more often unmarked than observers do. General knowledge differences should not contribute to this asymmetry, because the same information is expressed here in two different ways—with or without a mental state verb. However, a more specific cognitive mechanism governing belief reasons plays an important role: In their minds actors directly represent the content of their belief—for example, *the plants are dry*. They do not normally represent their own belief qua mental state; that is, they do not represent *I believe the plants are dry* (Moore, 1993; Rosenthal, 2005). As a result, when formulating their belief reasons in language, actors will typically describe what they represented and therefore leave their belief reasons unmarked: "Why did you turn the sprinkler on?"—"Because the plants were dry." Observers, by contrast, represent the actor *as having* certain beliefs—*she believed the plants were dry*—and they will tend to mark those beliefs with a mental state verb: "Perhaps she thought the plants were dry."

A second, more motivational process can contribute to the marker asymmetry as well, namely, the explainer's desire to convey an attitude toward the cited belief reason (Malle, 1999; Malle et al., 2000). Specifically, omitting a belief marker indicates the explainer's endorsement of that belief, whereas using a mental state marker distances the explainer from the belief. For example, if an explainer says, "She turned on the sprinkler because the plants were dry," the explainer himself seems to believe that the plants were dry. By contrast, if he says, "She turned on the sprinkler because *she thought* the plants were dry," he distances himself from the actor's belief. By explicitly stating that the actor thought the plants were dry, the explainer suggests that there is some doubt about the truth of the actor's belief. Actors can use this same linguistic device to distance themselves from their own past reasons ("I only locked the door because I thought you had already left"), but under normal circumstances they will be less likely than observers to make use of mental state markers as a distancing device.[2]

Methodological Approach

Most studies testing the traditional actor–observer asymmetry have used rating scales to assess how important each type of cause (e.g., person vs. situation) was in making the agent behave the way she did (Robins et al., 1996; Storms, 1973; Taylor & Fiske, 1975). Such scales have three chief disadvantages (Malle et al., 2000). First, they only weakly indicate what people actually do when they explain behavior, which is to express an explanation as a verbal statement (in private thought or conversation) that provides an answer to a why question (Hilton, 1990; Kidd & Amabile, 1981; Malle & Knobe, 1997b; Turnbull, 1986). Second, rating scales entail an a priori theoretical decision about what concepts people use in explaining behavior (White, 1993), thereby preventing the investigation of rival theoretical models. Instead of being forced to translate their explanations into theory-framed numerical ratings, participants can be asked to offer explanations in their own words. This methodology preserves the conceptual assumptions people themselves make in their behavior explanations and permits the coding of explanations in terms of competing theoretical models. Third, studies that used rating scales in the past have not provided evidence for the traditional actor–observer asymmetry, whereas studies using the free-response methodology have at least shown a tendency in that direction (Malle, 2006). The free-response approach thus provides the best possible tool to put the traditional asymmetry to the test.

In the present studies, free-response explanations were coded using the comprehensive *F.Ex* coding scheme (Malle, 1998/2007), which classifies explanations both in the terms of the folk-conceptual theory of explanation and in the terms of traditional attribution theory. It has been used in previous research projects (Dimdins, Montgomery, & Austers, 2005; Kiesler, Lee, & Kramer, in press; Knight & Rees, in press; Levi & Haslam, 2005; Malle, 1999; Malle et al., 2000; O'Laughlin & Malle, 2002) and has shown good reliability and predictive validity.

We aimed to design our studies in a cumulative manner, examining the predicted actor–observer asymmetries across a variety of contexts. Some studies asked people to recall behaviors over which they had puzzled along with the corresponding explanations that clarified them; others identified spontaneous explanations in conversation. Some let people choose the behaviors they explained; in others the experimenter selected those behaviors. All in all, we report on six studies of actor–observer asymmetries that tested the folk-conceptual and traditional attribution predictions about actor–observer asymmetries in explanation. The results of three additional studies are then included in a meta-analysis that provides a comprehensive test of the predictions.

Study 1

Method

Participants and procedure. Undergraduate students (N = 139) in an introductory psychology course completed the explanation measure as part of a survey packet during a group testing session in exchange for partial credit toward a course requirement. Debriefing was given for the survey packet as a whole.

Material. Participants were instructed to remember the last time they tried to explain (a) someone else's behavior or experience and (b) their own behavior or experience (in counterbalanced order; see Malle & Knobe, 1997b, Study 1, Form A). After describing each behavioral event, participants wrote down how they explained it.

Coding. All explanations were coded using the F.Ex coding scheme (Malle, 1998/2007). After training on 20 cases, two coders independently classified all remaining explanations. (For reliabilities see the Appendix.) The coding categories included the three

[2] One might think that a similar logic should hold for desire reasons as well. But it turns out that desire markers do not serve a distancing function (Malle et al., 2000). At least in English, the grammatical forms of marked and unmarked desires are highly similar ("Why are you rushing?"—"I want to be on time" vs. "To be on time"). Because of the lack of clear linguistic differentiation, mental state markers for desire reasons cannot indicate one's endorsing or distancing attitude toward the actor.

folk-conceptual parameters (reason vs. causal history explanations; belief vs. desire reasons;[3] marked vs. unmarked belief reasons) and several variants of the traditional attribution distinctions. Following instructions and examples in Nisbett et al. (1973), McGill (1989), and Shaver, Gartner, Crosby, Bakalarova, and Gatewood (2001), coders classified all explanations into person and situation attributions[4] and, within person attributions, into traits and nontraits. (Traits were defined as those stable dispositions that are attributes of one's personality.) Person versus situation was also separately classified within causes, causal histories, and reason contents; trait versus nontrait was separately classified within causes and causal histories.

Analyses. The dependent variables were raw counts of explanation parameters. That is, each participant was assigned a score for the number of reasons he or she gave from the actor perspective and a score for the number of reasons he or she gave from the observer perspective, and likewise for causal history explanations, belief reasons, desire reasons, and so on. However, even though many participants provided explanations from both perspectives, a given person often produced two explanations of different types (e.g., a person cause from the actor perspective and a reason from the observer perspective). As a result, the cell sizes were too small for within-subject tests of perspective. We therefore adopted a between-subjects approach that treats behavior as the unit of analysis. After separating each participant's data into an actor response vector and an observer response vector, we correlated the two vectors to examine the independence assumption. The vectors were indeed independent, with correlations for the explanation parameters averaging $r = .03$ ($SD = .10$).

All hypotheses were examined using mixed analysis of variance (ANOVA),[5] testing statistical interactions of the perspective factor with the relevant explanation parameter's within-subject factor. For the folk-conceptual hypotheses, these factors were reasons versus causal histories, belief versus desire reasons, and marked versus unmarked belief reasons. For the person–situation hypothesis, the main factor was an overall person–situation classification of all explanations, and follow-up analyses explored, where appropriate, which subtypes of explanations (for intentional or unintentional behavior) drove the effect. The trait hypothesis was tested as an unweighted average of separate tests for unintentional behaviors and intentional behaviors, averting data patterns known as Simpson's paradox (Simpson, 1951).[6] The separate tests are reported where appropriate.

Results

Analyses were based on a total of 438 explanations for 216 behaviors (110 intentional, 106 unintentional). Actor explanations constituted 51% of all explanations. Means are presented in Table 2; standard deviations are provided in the supplementary material. The results of the main hypothesis tests are displayed in Figure 1.

Folk-conceptual hypotheses. Within explanations for intentional behaviors, actors offered more reasons and fewer causal histories than observers did, $F(1, 109) = 10.4$, $p < .01$, $d = 0.61$, providing support for the reason asymmetry. Within reason explanations, actors offered more belief reasons and fewer desire reasons than observers did, $F(1, 69) = 6.5$, $p < .05$, $d = 0.60$, providing support for the belief asymmetry. Within belief reasons, actors used fewer mental state markers than observers did, $F(1, 32) = 10.1$, $p < .01$, $d = 1.20$, providing support for the marker asymmetry.

Traditional attribution hypotheses. Across all behavior explanations, the person–situation hypothesis was not supported ($F < 1$, $d = 0.11$), nor did we find any evidence when testing the hypothesis within explanation types (all Fs < 1). The trait hypothesis also received no support. Across all explanations observers offered no more traits than observers did ($F < 1$, $d = 0.13$), and neither of the specific explanation types showed a significant effect.

Discussion

Study 1 suggests that when explaining intentional actions, actors offer more reasons whereas observers prefer causal history explanations. When offering reason explanations, actors produce more belief reasons whereas observers tend toward desire reasons. And when specifically offering belief reasons, actors leave out the relevant mental state markers (e.g., "I know," "I thought") whereas observers often use such markers.

In testing multiple variants of the traditional actor–observer hypothesis, we found no evidence for any general person–situation difference and only a trend for observers to offer somewhat more traits when offering person causes. It should also be noted that trait explanations were quite rare. Eighty-seven of all participants mentioned no trait at all, and overall there were 0.27 trait explanations and 1.70 nontrait explanations per behavior explained. Thus, whereas the hypotheses derived from the folk-conceptual framework were confirmed, the person–situation hypothesis and the trait asymmetry were not.

This first study had several positive features: Participants recalled actual behaviors they had tried to explain; the explained behaviors covered the full range of intentional and

[3] We also coded a third reason type, *valuings*, but we had no interest in valuings per se, merely classifying them separately to provide a cleaner belief–desire test.

[4] We also coded interactions separately to provide a cleaner person–situation test.

[5] The dependent variables were counts, which could be considered a metric scale, but their distributions were frequently skewed. We also tested the five hypotheses using a conservative method of transforming each count into a dichotomous variable (0, 1) and conducting log-linear analyses between explanation parameters and the perspective factor. The results were exactly parallel to those from the ANOVA, confirming (or disconfirming, in some cases) the same hypotheses as the ANOVAs did. We therefore report only the more precise and statistically more powerful ANOVA results.

[6] The base rates of traits were higher for unintentional behaviors (12% traits per person for causes) than for intentional behaviors (29% traits per person for causal history factors). In addition, actors explained more unintentional behaviors whereas observers explained more intentional behaviors (and offered more causal history explanations for those than actors did). A standard (weighted) average opens the door to Simpson's (1951) paradox, as observers could appear to offer more traits merely because they explain more intentional behaviors and do so with more causal histories. In Study 5, for example, observers showed fewer trait explanations than actors within intentional behaviors ($d = -0.45$) and the same number as actors within unintentional behaviors ($d = 0.02$). A standard aggregate analysis would have portrayed observers as providing more trait explanations ($d = 0.15$); an unweighted average corrects this result ($d = -0.29$).

Table 2
Means for All Actor-Observer Hypotheses Across Six Studies

	Study 1		Study 2		Study 3		Study 4			Study 5				Study 6			
Explanation category	Actor	Obs	Actor	Obs	Actor	Obs	Actor	Distant obs	Close obs	Actor	Matched distant obs	Close obs	Matched distant obs	Actor	Matched obs	Matched IM obs	IM Actor
Explanation mode																	
Reasons	1.31	0.70	1.27	0.99	1.23	0.85	0.96	0.46	0.61	1.82	1.46	1.45	1.47	1.79	1.04	1.81	1.96
Causal history (CHR)	0.62	1.14	0.23	0.42	0.21	0.40	0.23	0.49	0.57	0.36	0.62	0.62	0.59	0.58	0.93	0.83	0.87
Reason types																	
Belief	0.90	0.32	0.70	0.49	0.91	0.57	0.59	0.27	0.34	0.97	0.54	0.98	0.60	1.53	0.83	1.12	1.27
Desire	0.56	0.74	0.51	0.59	0.31	0.48	0.60	0.66	0.74	0.90	1.05	0.60	0.78	0.39	0.52	0.76	0.59
Belief reasons																	
Unmarked	1.20	0.33	0.62	0.33	1.18	0.79	0.97	0.57	0.62	1.23	0.83	0.94	0.68	1.33	0.77	0.51	1.19
Marked	0.24	0.78	0.08	0.15	0.18	0.32	0.30	0.64	0.49	0.46	0.53	0.70	0.76	0.54	0.39	0.86	0.40
All person-situation																	
Person	1.42	1.55	2.39	2.82	0.72	0.67	1.10	1.02	1.20	1.47	1.40	1.24	1.42	1.25	1.13	1.78	1.62
Situation	0.26	0.24	1.31	0.71	0.51	0.43	0.25	0.09	0.10	0.52	0.45	0.42	0.32	0.61	0.53	0.46	0.76
Among CHRs																	
Person	1.20	1.59	0.57	1.10	0.62	0.81	0.79	1.00	1.00	1.00	1.34	0.96	0.91	0.73	0.61	0.89	0.46
Situation	0.05	0.18	0.29	0.09	0.39	0.29	0.13	0.09	0.11	0.25	0.24	0.37	0.18	0.38	0.62	0.20	0.11
Among causes																	
Person	1.46	1.50	1.03	0.88	0.73	0.63	0.96	1.08	0.91	1.52	1.21	1.26	1.32	1.26	1.23	1.58	1.86
Situation	0.23	0.34	0.40	0.31	0.48	0.45	0.18	0.00	0.14	0.54	0.67	0.36	0.59	0.47	0.39	0.59	0.81
Reason contents																	
Person	0.78	0.58	0.38	0.57	0.41	0.45	0.78	0.69	0.79	0.71	0.74	0.67	0.56	0.59	0.48	1.09	0.68
Situation	0.58	0.58	0.76	0.55	0.67	0.43	0.44	0.37	0.38	0.70	0.58	0.64	0.64	1.03	0.63	0.70	1.17
All traits																	
Trait	0.41	0.52	0.27	0.39	0.57	0.54	0.09	0.28	0.15	0.30	0.19	0.41	0.07	0.33	0.13	0.04	0.10
Nontrait	1.23	1.26	1.01	0.99	0.57	0.44	1.05	0.96	1.06	1.33	1.54	1.10	1.59	1.12	1.44	1.55	1.78
Among CHRs																	
Trait	0.54	0.47	0.50	0.71	0.00	0.42	0.12	0.46	0.17	0.50	0.30	0.55	0.09	0.25	0.17	0.00	0.13
Nontrait	1.06	1.33	0.55	0.67	1.14	0.96	0.98	0.73	1.07	0.83	1.28	0.80	1.46	0.88	1.17	1.25	1.25
Among causes																	
Trait	0.29	0.57	0.05	0.08	0.07	0.29	0.06	0.10	0.12	0.09	0.08	0.27	0.05	0.40	0.10	0.08	0.08
Nontrait	1.41	1.19	1.48	1.30	1.20	0.83	1.11	1.20	1.06	1.82	1.79	1.39	1.72	1.36	1.71	1.84	2.30

Note. Means are numbers of explanations of each category. Obs = observer; IM = impression-managing.

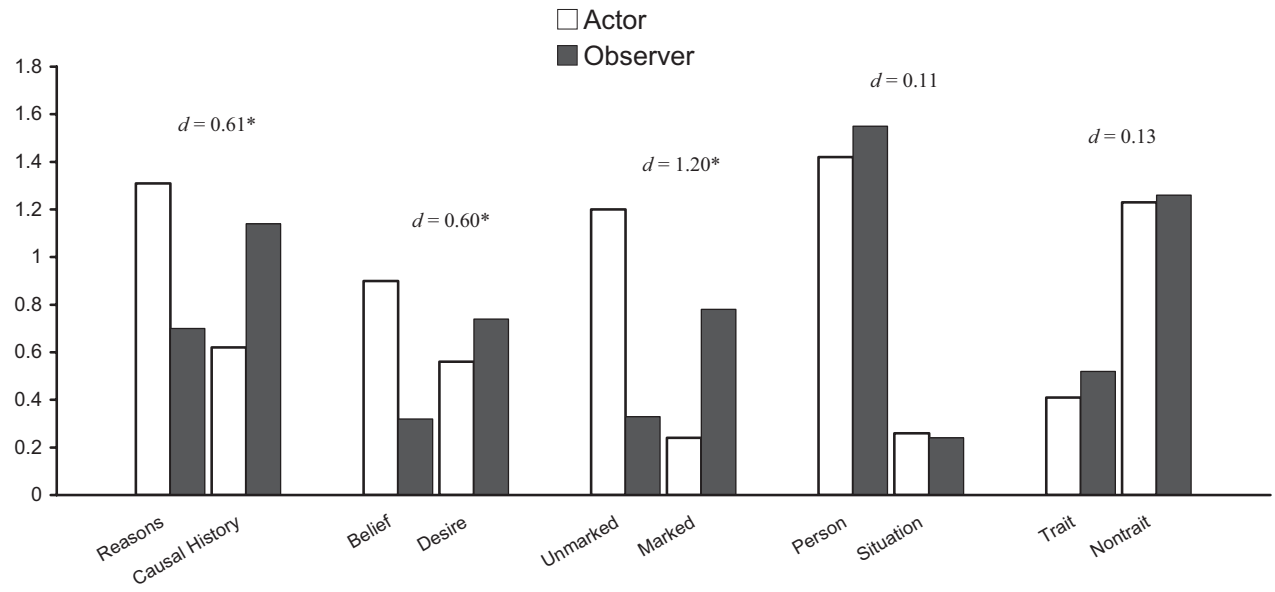

Figure 1. Actor–observer asymmetries tested in Study 1: three folk-conceptual hypotheses (reason asymmetry, belief asymmetry, and marker asymmetry) and two traditional hypotheses (person–situation asymmetry, trait asymmetry). Significant interaction effects are indicated by an asterisk.

unintentional as well as observable and unobservable events, distributed with the same frequencies in which they naturally occur (Malle & Knobe, 1997b); and the explanations covered the full range of explanation parameters, both for traditional attribution hypotheses and for the three folk-conceptual hypotheses. However, the study had problematic features as well. Participants used more causal history explanations overall than we had found in previous studies (cf. Malle, 1999; Malle et al., 2000), and so the effect sizes we computed here may not be entirely representative. More important, we had no experimental control over the behaviors people recalled and explained. It is possible that actors and observers appear to differ in their explanations because they select subtly different types of behaviors, and those behaviors then demand different explanations. In the next study we sought to replicate the findings of Study 1 but remedy this problem by controlling for the behaviors that actors and observers explain.

Study 2

Method

Participants and procedure. Introductory psychology students ($N = 221$) completed the explanation measure as part of a group-testing survey packet. They received partial credit toward a course requirement. Debriefing was given for the survey packet as a whole.

Material. Participants completed a one-page measure in which they explained three behaviors either from the actor perspective or from the observer perspective. On the basis of pretests, we selected stimulus behaviors that almost everyone had performed or had observed another person perform. We also tried to select socially relevant behaviors that would make it likely that participants cared about the behaviors and their explanations. Each measure contained three behavior types: one positive intentional, one negative intentional, and one negative unintentional. To increase generalizability, each of these types was represented by three specific behavior descriptions. The specific triplets of behaviors were counterbalanced across participants. Thus, a given measure contained one of the positive intentional behaviors ("went out of your way to help a friend"; "put a lot of energy into a class project"; "gave money or time to a charity"), one of the negative intentional behaviors ("insulted someone behind their back"; "teased somebody"; "told a lie"), and one of the unintentional behaviors ("started crying"; "accidentally missed an appointment"; "suddenly got really angry").

The instructions to the explanation measure read as follows:

> Below we ask you to remember a specific time when you [some other person] behaved a certain way. Once you are able to clearly picture this behavior in your mind, please answer the question "Why did you [the person] do that?," using simple, everyday terms.

To help participants picture each event, we asked them to indicate when the event occurred and, in the observer condition, who performed the behavior. Then they wrote down why they, or the other person, performed the behavior.

Design and analysis. Perspective was a between-subjects factor, and explanation parameters formed levels of within-subject factors (e.g., reason vs. causal history, belief vs. desire). Dependent variables were raw numbers of each explanation parameter, as described in Study 1.

Results

Analyses were based on a total of 824 explanations for 543 behaviors (366 intentional, 177 unintentional). Actor explanations constituted 51% of all explanations. Means are provided in Table 2.

Folk-conceptual hypotheses. Among explanations for intentional behaviors, actors offered more reasons and fewer causal histories than observers, $F(1, 217) = 16.8$, $p < .001$, $d = 0.55$. This interaction effect corroborates the reason asymmetry. When explaining behaviors with reasons, actors offered more belief reasons and fewer desire reasons than observers, $F(1, 208) = 5.2$, $p < .05$, $d = 0.31$, corroborating the belief asymmetry. When explaining behaviors with belief reasons, actors offered more unmarked beliefs and fewer marked beliefs than observers, $F(1, 136) = 7.5$, $p < .01$, $d = 0.47$, corroborating the marker asymmetry.

Traditional hypotheses. This time, the person–situation hypothesis was supported in the overall analysis. Across all explanations, actors made fewer person references and more situation references than observers, $F(1, 219) = 13.6$, $p < .001$, $d = 0.49$. Examining more specific person–situation effects, we found that actors and observers did not differ in cause explanations ($d = -0.05$), even though the negative valence of the unintentional behaviors in this study provided favorable conditions for finding the effect (Malle, 2006). There was an actor–observer difference in causal history explanations, $F(1, 86) = 13.9$, $p < .001$, $d = 0.80$. By contrast, the trait hypothesis was not supported. Actors cited slightly fewer traits than observers, but the difference was not significant, $F(1, 192) = 1.0$, $d = 0.15$.

Discussion

The hypotheses derived from the folk-conceptual framework were again supported, even when we controlled for the specific behaviors that actors and observers explained. Actors offered relatively more reasons than observers, more belief reasons, and more unmarked belief reasons. This study, unlike Study 1, provided some support for the person–situation asymmetry, but the effect was limited to explanations of intentional behavior. The trait hypothesis, however, received no support.

Study 3

Whereas we imposed more control on people's explanations in Study 2 than in Study 1, we went in the opposite direction with Study 3, opening the investigation to more naturally occurring behavior explanations. To collect a naturally occurring set of behavior explanations that were not elicited by a researcher, we extracted explanatory statements from conversations between pairs of participants.

Method

Participants and procedure. Seventy-six undergraduate students participated in the study. They received partial credit toward a course requirement and were debriefed at the end of the study. Seventy-one of the participants provided spontaneous behavior explanations. Each participant was paired with either a friend whom the person had brought along or a stranger (another undergraduate student). Each pair had two conversations of about 8 min each. In one conversation, Person A described an upsetting event to Person B; in the other, Person B described a confusing event to Person A. The assignment of events to persons was random, and the order of conversation topics was counterbalanced across subjects.

Extraction of explanations from conversations. A candidate explanation was transcribed from the audiotaped conversations if it contained the keywords *because*, *'cause*, *since*, *(in order) to*, or *so that* or when it constituted an answer to a why question. A reliability check on the identification of explanations from 10 conversations yielded 90% agreement among two coders. The remaining conversations were distributed among three coders, who individually analyzed them, but random checks were conducted for possible misses. Next, each of the 645 extracted passages was judged for being codable as a behavior explanation, with 89% coder agreement. After discussion of disagreements, the coders eliminated passages that did not represent clear behavior explanations. These passages included, out of the original number, 3% unclear or missing explanations, 5% nonbehavioral events, 6% ambiguous agents due to passive voice, and 8% claim backings (statements beginning with *because* that are not explanatory but provide evidence for a prediction or claim, e.g., "That's hard too, because that puts you more into the parenting role."). In addition, 2% of the behaviors were performed by group agents and were excluded because group explanations differ in systematic ways from individual explanations (O'Laughlin & Malle, 2002). Two coders then classified these explanations using the F.Ex coding scheme (Malle, 1998/2007). Reliabilities are shown in the Appendix.

Analyses. The results feature participants as units of analysis, with scores averaged across multiple explained behaviors per person. (Analyses using behaviors as units of analysis yielded highly similar results.) For the same reason as in Study 1, we treated the two data vectors of actor explanations and observer explanations as levels of a between-subjects factor as they were again independent, with correlations for the various explanation parameters averaging $r = .07$ ($SD = .09$). Moreover, those within-subject analyses that had acceptable cell sizes showed patterns of results that were highly similar to the ones reported below.

Results

The results were based on a total of 597 explanations for 449 behaviors (260 intentional, 189 unintentional). Actor explanations constituted 64% of all explanations. Means are presented in Table 2.

Audience. Some conversations occurred among strangers, others among relatively intimate friends. However, none of the hypothesis-relevant actor–observer patterns interacted with level of intimacy among conversation partners, and so analyses were collapsed over this factor.

Folk-conceptual hypotheses. Among explanations of intentional behavior, actors offered more reasons and fewer causal histories than observers, $F(1, 86) = 10.6$, $p < .01$, $d = 0.69$, supporting the reason asymmetry. When using reason explanations, actors offered more beliefs and fewer desires than observers, $F(1, 78) = 4.8$, $p < .05$, $d = 0.49$, supporting the belief asymmetry. When citing belief reasons, actors offered more unmarked beliefs and fewer marked beliefs than observers, $F(1, 54) = 4.0$, $p < .05$, $d = 0.54$, supporting the marker asymmetry.

Traditional hypotheses. The person–situation hypothesis was not supported ($F < 1$, $d = 0.03$). By contrast, the trait hypothesis was supported. When providing person attributions, actors offered

fewer traits and more nontraits than observers, $F(1, 80) = 15.1$, $p < .001$, $d = 0.88$, within both cause explanations ($d = 0.85$) and causal history explanations ($d = 0.76$).

Discussion

This study examined spontaneous explanations people offered during conversation. Even in this unstructured social context, all three folk-conceptual hypotheses were replicated, with effect sizes between 0.49 and 0.69. Among the traditional hypotheses, the person–situation asymmetry failed to hold, casting some doubt on the effect found in Study 2. The trait asymmetry emerged this time, for both intentional and unintentional behaviors.

Whether people explained behaviors to a friend or a stranger did not moderate any of these results. Thus, the processes that drive actor–observer asymmetries do not include the explainer's degree of intimacy with the person *to whom* a behavior is explained. However, the level of intimacy between the observer and the actor *whose* behavior is explained may well moderate actor–observer asymmetries. In fact, considerations of this variable led to the knowledge account of the traditional attribution asymmetry (Jones & Nisbett, 1972), often featured in textbooks (e.g., Bernstein, Clarke-Stewart, Roy, & Wickens, 1997; Franzoi, 2006; Gray, 2002; Meyers, 2004; Taylor, Peplau, & Sears, 2006).

According to the traditional knowledge account, observers normally lack intimate or privileged knowledge about the actor (e.g., feelings, intentions, and personal history) and are therefore less able to provide situational explanations of the actor's behavior. Observers who are on more intimate terms with the actor should therefore increase their situational attributions. In the folk-conceptual model, too, information access is postulated to be one of the driving forces of explanatory asymmetries, as analyzed earlier. Study 4 therefore examined the role of knowledge on both sets of predicted actor–observer asymmetries.

Study 4

The established approach to testing the knowledge account is to compare two types of observers' explanations of an actor's behavior: distant (stranger) and close (intimate). This approach presumes knowledge to be a relatively stable cognitive structure that is acquired through relationships over time. Another way to conceive of knowledge is as a temporary resource that explainers can acquire in a specific context: knowledge about one particular behavior that another person performed. This sort of knowledge could be available, for example, when the observer is copresent with the actor, directly observing the behavior in question. In this study we examined both stable knowledge and copresence and their independent effects on each of the predicted actor–observer asymmetries (reason vs. causal history, trait vs. nontrait, etc.).

Method

Participants and procedure. Of 416 undergraduate students in an introductory psychology course who were given a group-testing survey packet, 398 completed the relevant explanation measure. Participants were debriefed at the end of the testing session and received partial credit toward a course requirement.

Material. The explanation measure consisted of a one-page questionnaire that elicited (in counterbalanced order) one behavior and its explanation from the actor perspective and one behavior and its explanation from the observer perspective. For actor explanations participants were asked to "recall the last time YOU performed an action that another person (other people) found surprising." Observer explanations fell into one of four between-subjects conditions, produced by crossing knowledge (close vs. distant observer) with copresence (yes vs. no). We manipulated copresence by asking participants to "recall the last time you SAW a stranger perform a puzzling action" (copresence) or "recall the last time you HEARD about a stranger performing a puzzling action" (no copresence). We manipulated knowledge by asking participants to explain the action of "someone you know well" or the action of "a stranger." Participants first described the action and then answered two manipulation check questions (with response options in parentheses): "Were you present when the action occurred?" (*yes*, *no*) and "How well do you know the person?" (3-point rating from *barely* to *very well*). Explanations were F.Ex-coded as in Studies 1–3, and reliabilities are documented in the Appendix.

Analyses. As in previous studies, we separated actor responses and observer responses from the same participants to allow for a between-subjects analysis. The intercorrelations of actor and observer responses for the various explanation parameters supported the independence assumption, averaging $r = .02$ ($SD = .19$).

The factor of explainer role (actor, close observer, distant observer) was divided into two orthogonal contrasts. The first compared actor explanations with distant observer (stranger) explanations, attempting to replicate the actor–observer asymmetries from Studies 1–3. The second compared distant observers with close observers and thus examined the knowledge hypothesis. Both contrasts were tested against the same overall error term. The copresence hypothesis was tested by comparing distant observers who were present with distant observers who were absent. (This comparison is less meaningful for close observers because in that case knowledge and copresence are confounded.)

Manipulation checks. The knowledge manipulation had its expected effect on the subjective knowledge ratings ($M = 2.6$ for close observers and $M = 1.2$ for distant observers). The copresence manipulation also had its expected effect, with 84% of those in the copresence condition reporting that they were actually present when the action happened, compared with 29% in the other absent condition. (Analyses with and without the participants who did not respond as expected showed the same results.)

Results

Analyses were based on a total of 1,014 explanations for 709 behaviors (77 unintentional). Means are presented in Table 2.

Supporting the reason asymmetry, actors offered more reasons and fewer causal histories than distant observers did, $F(1, 792) = 6.7$, $p < .01$, $d = 0.79$. The corresponding knowledge hypothesis was not confirmed, as close observers were indistinguishable from distant observers, $F(1, 792) < 1$, $d = 0.08$. The copresence hypothesis was also not confirmed, as copresent and absent observers were indistinguishable from each other, $F(1, 541) < 1$, $d = 0.08$.

Actors also offered more beliefs and fewer desires than observers did, $F(1, 464) = 6.0$, $p < .05$, $d = 0.33$, supporting the belief asymmetry. The corresponding knowledge hypothesis was not confirmed, as close observers were indistinguishable from distant observers, $F(1, 463) < 1$, $d = -0.02$. The copresence hypothesis was also not confirmed, as copresent observers were indistinguishable from absent observers, $F(1, 343) < 1$, $d = 0.04$.

The marker asymmetry was replicated, $F(1, 180) = 6.0$, $p < .05$, $d = 0.69$, and once again, the knowledge hypothesis was not confirmed, $F(1, 180) < 1$, $d = 0.18$. This time, however, the copresence hypothesis was supported, $F(1, 143) = 5.04$, $p < .05$, $d = 1.23$, so much so that copresent observers favored unmarked belief reasons exactly as much as actors did ($M_{diff} = 0.67$), whereas absent observers clearly favored marked belief reasons ($M_{diff} = -0.63$). This finding must be treated with caution, however, as only 14 distant observers were compared with 132 actors.

Traditional asymmetries. The overall person–situation asymmetry was not confirmed, $F(1, 697) < 1$, $d = 0.09$. A small knowledge difference pointed in the opposite direction from what traditional theory would have predicted. Close observers offered slightly more person attributions and slightly fewer situation attributions than distant observers, $F(1, 697) = 2.7$, $p < .10$, $d = -0.21$. No copresence effect emerged ($F < 1$). The trait asymmetry received some support. When offering person attributions, actors cited fewer traits and more nontraits than observers did, $F(1, 251) = 3.9$, $p < .05$, $d = 0.33$. The breakdown into intentional and unintentional behaviors showed that observers used more causal history traits than actors did, $F(1, 188) = 11.3$ $p < .001$, $d = 0.62$, but actors and observers used equal numbers of cause traits ($d = -0.06$, $F < 1$). There was also a small knowledge effect, as close observers offered fewer traits than distant observers did, $F(1, 251) = 4.0$, $p < .05$, $d = 0.26$. This effect, too, was visible for causal history traits ($d = 0.66$) but not for cause traits ($d = -0.18$). No copresence effect emerged ($F < 1$).

Summary. The three folk-conceptual hypotheses were again replicated, but knowledge had no effect on actor–observer asymmetries. Copresence had an effect specifically on the marker asymmetry. The person–situation asymmetry was again absent. A trait asymmetry was found this time, though only for traits in causal history explanations, which runs counter to Study 2, in which it was cause explanations that displayed a trait asymmetry. The knowledge hypothesis was supported only for the trait asymmetry in causal history explanations.

Follow-up study. Surprised by the paucity of knowledge differences, we conducted a follow-up study in which we tested solely the knowledge hypothesis and increased the representativeness of behaviors by asking undergraduate students to recall and explain five behaviors performed by strangers (distant observer condition) and five behaviors performed by friends or family members (close observer condition). Fifty-six participants explained 506 behaviors with 1,008 explanations, which were coded, aggregated per person, and analyzed with observer type as a within-subject factor. The results confirmed Study 4's general lack of knowledge effects. In fact, the correlation between relevant effect sizes of these two studies across the various explanation parameters was $r(10) = .69$. For the folk-conceptual parameters, the knowledge effect sizes ranged from −0.24 to −0.08. For the person–situation comparisons, the effect sizes ranged from −0.26 to 0.24. Only the trait parameters showed a significant knowledge effect (as in Study 4): $d = 0.37$ for traits overall ($p < .05$), which was driven more by causal history traits ($d = 0.48$) than by cause traits ($d = 0.24$).

Discussion

In light of these findings, one might abandon the knowledge hypothesis for all but one explanation parameter: that of traits in causal history explanations. However, one central feature of Study 4 (and its follow-up) may have made the detection of knowledge effects overly difficult: Both close and distant observers self-selected the behaviors they explained. When given this chance to self-select, people will tend to choose behaviors that they can explain reasonably well. Thus, distant observers, who may normally be at a disadvantage when explaining other people's behavior, can overcome this disadvantage by suitably choosing behaviors that they find easy to explain. As a result, any naturally occurring knowledge differences between close and distant observers would be difficult to detect. Study 5 therefore required distant observers to explain behaviors that they had not themselves selected.

Study 5

While gaining control over the behaviors that both types of observers explained, this study also increased the realism of the behaviors in question. We asked participants to describe a conflict they had had with another person. Then the experimenter selected 8–10 behavioral events from this audio-recorded description and asked participants to explain each event—from the actor perspective for behaviors they had performed themselves and from the close observer perspective for behaviors that their conflict partners had performed. Distant observers were recruited in a second sample and were each matched to one participant from the original sample. They listened to their matched participant's original audio-recorded conflict description and explained the same behaviors that the initial participant had explained. This way, close observers explained behaviors that actually occurred in a situation of intimate contact, and distant observers were required to explain those same behaviors.

Method

Participants. Fifty undergraduate students constituted the first sample, which provided actor explanations and close observer explanations. A second sample of 50 students constituted the distant observers, who were matched in pairwise fashion to the original participants. Four matches could not be achieved (two distant observers did not offer any codable explanations; two original conflict recordings had been damaged). All participants received partial credit toward a course requirement and were debriefed at the end of the study.

Procedure. Initial participants were asked to describe "the last time you had an interesting conflict with a romantic partner, friend, or parent." This description was audio-taped. While participants were occupied with another task, one experimenter listened to the tape and selected 8–10 behavioral events from the conflict description. The experimenter attempted to select events that were not already explained by the speaker and sought a balance between

positive and negative events, between intentional and unintentional events, and between actor and observer events. A different experimenter then asked participants to explain each of the selected behavioral events either in a questionnaire (written format) or in an audiotaped interview with the experimenter (spoken format). (Format had no impact on the results and is not further discussed.) This procedure elicited explanations from the actor perspective (participants explaining their own behavioral events) and from the close observer perspective (participants explaining their conflict partners' behavioral events). Two coders (interrater agreement, $\kappa > .90$) rated the intimacy of close observers on a scale from 0 (*strangers*) to 3 (*close relatives, close friends, and romantic partners*). Close observers were judged to have an average intimacy of 2.5 with the agents whose behavior they explained.

Each participant in the second, distant observer sample was matched with one of the original participants. Each new participant listened to the initial participant's audiotaped conflict description and explained exactly the same behaviors that the original participant had explained, formulated in the third person. This way, 46 matched pairs were formed.

Material. Each of the 8–10 selected behavioral events was restated and followed by a why question to elicit participants' explanations. For example, an item eliciting an actor explanation was "You said: 'I felt guilty for being here.' Why did you feel that way?" The corresponding distant observer's item was "She said: 'I felt guilty for being here.' Why did she feel that way?"). All explanations were F.Ex-coded as in Studies 1–4.

Analysis. With participants as units of analysis, explanation parameters (e.g., reasons) were averaged across explained behaviors (e.g., three intentional behaviors) within the actor, close observer, and distant observer perspectives. Actors and their matched distant observers explained the same behaviors, permitting a repeated measures test of all actor–observer asymmetries. Similarly, close observers and their distant observers explained the same behaviors, permitting a repeated measures test of the knowledge hypothesis. Actors and close observers were not directly compared because they explained different behaviors. Analyses were conducted on 1,591 explanations for 578 behaviors (334 intentional, 244 unintentional).

Results

Folk-conceptual hypotheses. Supporting the reason asymmetry, actors offered more reasons and fewer causal histories than their matched distant observers, $F(1, 39) = 6.5$, $p < .05$, $d = 0.57$. There was no corresponding knowledge effect, as close observers and their matched distant observers did not differ, $F(1, 39) < 1$, $d = -0.04$. Supporting the belief asymmetry, actors offered more beliefs and fewer desires than distant observers, $F(1, 36) = 6.1$, $p < .05$, $d = 0.49$. There was a strong knowledge effect, as close observers offered more beliefs and fewer desires than distant observers, $F(1, 35) = 8.3$, $p < .01$, $d = 0.52$. The marker asymmetry was in the predicted direction, as actors offered more unmarked beliefs and fewer marked beliefs than distant observers, but within this small sample the effect was not reliable, $F(1, 19) = 1.9$, $p = .19$, $d = 0.41$. There was no reliable knowledge effect, $F(1, 24) < 1$, $d = 0.24$.

Traditional hypotheses. No overall person–situation asymmetry emerged, $F(1, 44) < 1$, $d = 0.00$. A trend of a knowledge effect emerged such that close observers referred to somewhat fewer person factors and more situation factors than distant observers, $F(1, 45) = 4.0$, $p = .11$, $d = 0.26$, but this pattern was not robust across person–situation comparisons within causes, causal histories, and reason contents (ds $= -0.11$ to 0.09). The trait hypothesis was not supported, and means actually went in the opposite direction, $F(1, 39) = 1.7$, *ns*, $d = -0.29$. However, a strong knowledge effect counter to traditional predictions emerged, as close observers offered more trait explanations than distant observers, $F(1, 26) = 6.3$, $p < .05$, $d = -0.65$. This pattern held within both cause explanations ($d = -0.46$) and causal history explanations ($d = -0.78$).

Discussion

This study once again tested the knowledge hypothesis of actor–observer asymmetries in attribution, according to which close (intimate, familiar) observers show smaller asymmetries than do distant observers. The study design aimed at more realism for the behaviors explained and the context of providing explanations and, most important, held constant the specific behaviors that close and distant observers explained. This cleaner test of the knowledge hypothesis offered two noteworthy findings: a substantial knowledge effect for the belief asymmetry and a reverse knowledge effect for trait explanations. However, because of the prominence that the knowledge account has in the literature, we discuss all asymmetries in turn.

Reason asymmetry. Actors offered significantly more reasons and fewer causal history explanations than both close and distant observers. This finding, consistent with Study 4 and its follow-up, suggests that the reason asymmetry is not driven by general knowledge differences. At first glance, this may seem surprising, because reasons would be considered privileged knowledge to which an intimate observer should have relatively more access. However, many causal history explanations are privileged as well, referring to the agent's personality, past experiences, or unconscious motives. Intimate observers may have equal knowledge gains about the agent's causal history of reasons and about the reasons themselves; general knowledge therefore does not alter the balance between reason explanations and causal history explanations.

A more specific information access mechanism, however, is likely to influence the actor–observer asymmetry, namely, actors' ability to directly recall their own reasons (which they considered during deliberation), compared with observers, who must guess them or infer them from observable sources. Especially for actions that the actor performed after some deliberation, recalling a substantial number of reasons will be easy, and this memory advantage should contribute to a reason asymmetry. We are currently testing this hypothesis in our laboratory.

Belief asymmetry. The belief asymmetry was replicated once more with distant observers, but close observers showed a different pattern, mimicking actors, who offered more belief reasons than desire reasons. This pattern and its effect size of 0.52 suggest that a lack of agent-specific knowledge normally makes a strong contribution to the belief asymmetry. Conversely, knowing more about the agent, the action, and its context illuminates the agent's subjective beliefs—regarding details of the considered actions, their potential consequences, and facilitating or hindering features

of the context. Consider the following example, in which the close observer knows why the agent acted whereas the corresponding distant observer does not:

> (9) "She said: 'No, it was going to be two nights [of staying at the speaker's house].'" (Tell us why she said that.)
>
> Close observer: "Because *she thought it would be more convenient to stay another night* [marked belief] *since she had plans Saturday morning with her friend* [unmarked belief]."
> Distant observer: "*She was just letting him know that it was going to be two nights* [unmarked desire]."

We now have to clarify why close observers in Study 4 (and its follow-up) did not substantially increase belief reasons whereas those in Study 5 did. Our initial explanation was that the methodology of the previous studies, allowing explainers to choose their own behaviors, eliminated a real knowledge effect because distant observers were able to make up for their natural lack of information by selecting behaviors for which they had sufficient information. In Study 5, this option was not available (distant observers had to explain preselected behaviors), so the true informational difference between close and distant observers came to the fore.

A second, complementary explanation focuses less on distant observers being able to make up for their disadvantage than on close observers making use of their advantage. Richer knowledge about the agent, action, or context gives the close observer options to portray the agent in a more or less positive light. And just as actors increase belief reasons when they try to make themselves look good (Malle et al., 2000), observers may do the same when they offer charitable explanations of the actor's behavior. According to this account, close observers in Study 5 were more motivated than those in Study 4 to use their knowledge to portray familiar actors in a positive light. There is some auxiliary evidence in our data to support this assumption, namely, in the social desirability of the behaviors that each group selected.

In Study 4, both close and distant observers selected substantially less positive behaviors ($M = 0.59$) than actors did ($M = 1.31$), $p < .001$, $d = 0.63$, and close observers were indistinguishable from distant observers ($d = 0.04$). Thus, if the evaluative stance expressed in behavior selection is an indicator of impression management motivation, then both observer groups showed very little such motivation in Study 4, and fittingly they both used more desires and fewer beliefs than actors did. In the follow-up study (which contained no actor data), close and distant observers were again indistinguishable in their behavior selections ($d = 0.07$) and in their use of belief versus desire reasons. In Study 5, the average social desirability of behaviors was very similar for actors and close observers ($d = 0.17$, *ns*) and their belief rates were also similar (but distinct from those of distant observers), illustrating a knowledge effect. Thus, in Study 5, both in evaluative stance and belief rates, close observers looked quite like actors (who arguably have impression management motivation), whereas in Study 4 and its follow-up, close observers looked like distant observers (and neither of the observer groups showed much impression management). This indirect evidence suggests that close observers can overcome the belief asymmetry only when both intimate knowledge and the motivation to portray the actor in a positive light coincide. This motivation is examined more directly in Study 6.

Marker asymmetry. The marker asymmetry in Study 5 was in the predicted direction but did not reach statistical significance. However, none of the means of either actors or observers was an outlier relative to previous studies, and the effect size ($d = 0.41$) was still respectable. We can therefore assume that random variation and lower statistical power accounts for this result. Meta-analytic results reported later will confirm this assumption.

Traditional hypotheses. Both the person–situation asymmetry and the trait asymmetry failed to replicate in Study 5, whereas we saw a knowledge effect for the person–situation asymmetry and a reverse effect for the trait asymmetry. Neither of these knowledge patterns held in Study 4 and its follow-up. Thus, both the traditional actor–observer asymmetries themselves and their presumed moderator effects due to knowledge are highly inconsistent and resist further interpretation.

We now turn to a second possible determinant of actor–observer asymmetries in explanation: the explainer's motivation to manage the impression the agent (self or other) creates in an audience. Thus far, the evidence for such a process has been only indirect, and so we decided to examine impression management directly. Because intimacy and impression management may often be confounded (as Studies 4 and 5 suggested), we manipulated impression management in strangers, hence in the absence of any potential effects of intimate knowledge.

Study 6

Behavior explanations have a dual nature. They are not only a cognitive activity to find meaning in the world; they are a social activity to manage ongoing interactions (Malle, 2004). Explanations can be used to clarify, justify, defend, attack, or flatter; they serve as tools to guide and influence one's audience's impressions, reactions, and actions (Antaki, 1994; Goffman, 1959; Scott & Lyman, 1968; Semin & Manstead, 1983; Tedeschi & Reiss, 1981). Such impression management can be used from both the actor perspective and the observer perspective, but actors will more often portray themselves in a positive light. Thus, actors' greater use of impression management may help explain at least some of the actor–observer asymmetries we have documented in this article.

With respect to the reason asymmetry, one study showed that actors who had been invited to portray themselves as rational when offering behavior explanations to an audience significantly increased their use of reason explanations (Malle et al., 2000). What has not been tested is whether a generally positive portrayal operates much like a rational portrayal and whether observers, too, will provide more reason explanations when presenting the actor in a positive light. With respect to the belief asymmetry, our interpretation of Studies 4 and 5 suggested that knowledge by itself is not sufficient for an increase of observers' use of belief reasons; observers also have to be motivated to make the actor look good. But is such an impression management motive sufficient? Study 6 separates the potential role of impression management from that of knowledge by manipulating distant observers' attempts to make the actor look good.

It is somewhat unclear whether traditional actor–observer asymmetries were hypothesized to be subject to impression management motives. Jones and Nisbett (1972) and Nisbett et al. (1973) proposed that actors may try to protect their sense of freedom by favoring situational over dispositional attributions. There are conceptual problems with this proposal (Knobe & Malle, 2002), and the data have not been very supportive (e.g., Miller & Norman,

1975). But the traditional expectation may be that observers increase their situational attributions for actors whom they try to portray in a positive light.

Method

Participants. Undergraduate students participated for partial credit toward a course requirement and were debriefed at the end of the study. They were run individually but analyzed in matched pairs of actors and observers who explained the same behaviors. Of 62 participants, 4 (2 actors and their 2 matched observers) were excluded because no intentional behaviors had been selected. All analyses were therefore performed on 29 pairs of actors and observers, 15 in the control condition and 14 in the impression management condition.

Procedure. Participants were randomly assigned to one of four cells in a perspective (actor vs. observer) by motivation (impression management vs. control) design. In the actor condition, participants were audio-recorded telling a story from their personal life. While participants completed a few personality measures, the experimenter left the room, listened to the audio recording, and selected up to six behaviors that the participant had explicitly mentioned as having performed him- or herself in the story. The goal was to select at least three intentional behaviors and at least two unintentional behaviors (or experiences), both positive and negative, and formulate each of them as a why question (e.g., "Why did you go to Hawaii for the exchange program?"; "Why were you nervous about meeting new people?"). After completing the selection, the experimenter provided instructions to manipulate the participant's motivation:

> [All participants:] We have selected a few questions based on your story, and my research assistant will now ask you these questions. Please answer the questions as accurately as possible and as best as you can remember, but please keep your sentences short. Two to three sentences should be the average length. The questions will be about what you thought, felt, behaved, etc. [Only impression management:] Now here is the key point. Your goal when answering these questions is to create a positive impression. You want my research assistant to perceive you in as positive a light as possible. You do not need to lie in order to accomplish this, but rather phrase your answers in such a way that allows you to create a positive impression of yourself.

In the observer condition, each participant listened to a previously recorded actor's story and answered the same why questions that the matched actor had answered, assigned to the same experimental condition. The critical instruction, adjusted for the observer perspective, was "Your goal when answering these questions is to create a positive impression of this person. You want my research assistant to perceive this person in as positive a light as possible."

The research assistant, who was blind to the impression management manipulation, took the experimenter's place in the laboratory room and posed the preselected why questions. Participants' explanations were audio-recorded, transcribed, and F.Ex-coded as in Studies 1–5.

Results

Analyses were based on 692 explanations for 270 behaviors (117 unintentional). In a matched-pairs design, data from 29 actors and their corresponding observers (who explained the same behaviors) were treated as repeated measures. Fifteen pairs were in the control condition, 14, in the impression management condition. The latter group (both actors and observers) produced significantly more explanations overall ($M = 2.2$) than did the control group ($M = 1.5$).

We tested two main hypotheses: (a) the baseline actor–observer asymmetries in the control condition and (b) any differences between impression-managing observers and control observers (impression management hypothesis). To gauge the actor–observer asymmetry that would hold between impression-managing observers and control actors (who were not paired up in this study), we report its estimated effect size. In addition, (c) we note any differences between impression management actors and control actors. All means are displayed in Table 2.

Reason asymmetry. (a) The control condition showed the usual asymmetry, with actors offering more reasons (relative to causal histories) than observers did, $d = 1.03$, $F(1, 27) = 4.7$, $p < .05$. (b) When observers were instructed to make the agent look good, they offered noticeably more reasons than control observers did, $d = 0.71$, $F(1, 27) = 3.8$, $p = .06$. As a result, the reason asymmetry for impression-managing observers and control actors ($d = 0.20$) was one fifth of the baseline asymmetry's size ($d = 1.03$). (c) By contrast, actors instructed to make themselves look good did not differ significantly from control actors ($d = -0.10$). They even showed a tendency to offer more causal history explanations than control actors did, which runs counter to the idea that impression management is the primary cause of actors' substantial number of reason explanations. Instead, it is observers' apparent lack of impression management motives under normal circumstances that fosters a reason asymmetry.

Belief asymmetry. (a) The control condition showed the predicted asymmetry, with actors offering more belief reasons (relative to desire reasons) than observers did, $d = 0.71$, $F(1, 27) = 6.4$, $p < .05$. (b) When observers were instructed to make the agent look good, they increased both belief reasons and desire reasons, and so no difference in the relative importance of beliefs and desires emerged when compared with control observers ($F < 1$, $d = 0.07$). As a result, the belief asymmetry was not altered by impression management ($d = 0.66$). (c) Actors who were instructed to make themselves look good differed somewhat from control actors ($d = 0.39$), though not significantly so. The direction of this difference again ran counter to an impression management account, as impression-managing actors offered slightly fewer belief reasons (and more desire reasons). The standard belief asymmetry therefore does not appear to be simply a function of actors' impression management.

Marker asymmetry. (a) The control condition showed a noteworthy asymmetry in the usual direction ($d = 0.42$), though it was not statistically significant, $F(1, 23) = 1.1$. (b) When observers were instructed to make the agent look good, they differed from control observers by shifting from more unmarked beliefs (0.77 vs. 0.39) to more marked beliefs (0.51 vs. 0.86), $d = 0.80$, $F(1, 23) = 4.2$, $p = .05$. As a result, a substantial marker asymmetry emerged between impression-managing observers and control actors, $d = 1.18$, $F(1, 23) = 7.6$, $p = .01$, speaking against an impression management account of the marker asymmetry. (c) Actors who were instructed to make themselves look good showed the exact

same pattern as control actors ($d = 0.00$), both substantially favoring unmarked belief reasons.

Person–situation hypothesis. (a) No actor–observer asymmetry for person versus situation explanations emerged in the control condition, either in the overall classification ($d = -0.05$) or in any of the subsets (all Fs < 1). (b) Against traditional expectation, observers who tried to make the actor look good provided more person explanations than control observers did, $d = -0.69$, $F(1, 29) = 3.9$, $p = .06$. As a result, only impression-managing observers provided more person explanations than control actors did ($d = 0.61$). The absence of an actor–observer asymmetry in the control condition and the presence of such an asymmetry in the impression management condition is difficult to explain from an attribution standpoint. (c) Actors instructed to engage in impression management did not differ notably from control actors ($F < 1$); if anything, they increased overall person explanations as well ($d = -0.26$).

Trait hypothesis. (a) The actor–observer asymmetry for traits in the control condition pointed in the opposite direction to the traditional hypothesis, as observers actually used fewer traits than actors did ($d = -0.43$, *ns*). (b) Impression-managing observers used fewer traits yet, and so the reverse asymmetry between impression-managing observers and control actors bordered on traditional significance, $d = -0.65$, $F(1, 32) = 3.6$, $p < .10$. (c) Actors, too, decreased their use of traits in the impression management condition ($d = 0.80$), and reliably so, $F(1, 23) = 5.5$, $p < .05$.

Discussion

Tests of the three folk-conceptual asymmetries in Study 6 showed effect sizes comparable to the previous five studies, adding evidence to the replicability and stability of these asymmetries. By contrast, tests of the traditional attribution asymmetries were again not supported, further casting doubt on their strength and validity. Impression management motives on the part of observers specifically moderated the reason asymmetry but not the belief asymmetry or marker asymmetry. We discuss these results in turn.

Reason asymmetry. Observers who were motivated to portray the actor in a positive light produced almost as many reasons as actors themselves did, but without decreasing their causal history explanations. This pattern suggests that observers normally fail to offer reason explanations that they, in principle, could produce if only they made the effort. All observers were strangers to these actors and had little idiosyncratic information about them, and so the extra effort exerted by impression-managing observers must lie in attempts to take the actors' subjective perspective and infer or construct their idiosyncratic reasons for the particular action in the particular context. To illustrate, a count of the rare but telling linguistic expressions of inference ("I guess," "I think," "maybe," "probably") shows that impression-managing observers offered 18 explanations of intentional actions that contained an inference marker, compared with 11 among control observers. More important, 13 out of these 18 explanations by impression-managing observers were reasons, compared with 4 out of 11 for control observers.

This interpretation leads to the prediction that observers' rates of reason explanations should increase in response to direct perspective-taking instructions. Previous research in the context of traditional attribution theory used an "empathy" instruction and suggested that empathic observers provide more "situation attributions" for another person's behavior (Galper, 1976; Gould & Sigall, 1977; Regan & Totten, 1975). It is unclear how these findings translate into the effect of perspective taking on reason explanations, because the category of situation attributions can refer to a variety of different parameters distinguished by the folk-conceptual model of explanation—situational causes, situational causal history factors, or reasons with situation content. If the findings of Study 6 are a suitable indication, genuine situational causes and causal histories do not increase and may even decrease for impression-managing and perspective-taking observers. Perspective taking should, however, increase the use of reason explanations. Further, to the extent that the dominant class of reasons is unmarked beliefs, which typically have situation content (Malle, 1999), the seeming increase in situation attributions following empathy instructions in the literature may have resulted solely from explainers' consideration of reasons and the frequent situational content they represent.

Belief asymmetry. Impression management motives did not affect the actor–observer asymmetry for belief reasons. At first blush, this might seem to contradict the interpretation of Study 5, in which we argued that close observers overcame the actor–observer asymmetry because they cared to portray the actor in a positive light. But there we proposed that this elimination of the belief asymmetry requires two processes: first, the motivation to portray the actor in a positive light and, second, the availability of intimate knowledge. Study 4, we suggested, featured observers who had the knowledge but not the motivation; Study 6 featured observers who had the motivation but not the knowledge; and only Study 5 featured observers who had both—and that was the only time we saw the belief asymmetry eliminated. Naturally, this interpretation must be tested in future experiments that manipulate the two processes within the same sample.

Marker asymmetry. The actor–observer asymmetry of mental state markers for belief reasons was also unaffected by a general impression management motive. We continue to assume that belief markers serve the specific motivational goal of distancing oneself from an actor's belief reason ("She refused dessert because she thinks she's been gaining weight"; Malle et al., 2000). But this marker use does not necessarily make the *agent* look good (after all, it points out that the agent may be wrong); rather, it lets the *explainer* show that he knows better.

One question raised by Study 6 is why impression management instructions barely altered actors' patterns of explaining intentional behavior. In Malle et al. (2000), for example, actors who tried to make themselves look rational increased their belief reasons compared with control actors. However, when we compare the rates for belief and desire reasons in Study 6 with the corresponding numbers in Malle et al. (2000, Table 3), it appears that Study 6 contained a ceiling effect. In the 2000 study, control actors offered 0.8 belief and 0.5 desire, and actors in the rational self-presentation condition offered 1.1 beliefs and 0.4 desires. In the present Study 6, actors offered 1.5 beliefs and 0.4 desires in the control condition, which may be close to the ceiling for the number of belief reasons one can give without sounding unnatural. Even actors who are specifically instructed to make themselves look good cannot go beyond this ceiling. Thus, impression management may still contribute to the actor's side of the belief asymmetry; however, to

demonstrate such a contribution one may have to specifically restrict the actor's self-presentational behavior, which is not an easy feat.

In sum, the present results lend support to an impression management account of the reason asymmetry but not of any other actor–observer asymmetry. The data suggest that observers normally tend to withhold reason explanations but that when motivated to present the agent in a positive light, they actively infer and construct such reasons. This motivational process is likely to work alongside the basic process of information access. Observers may normally have difficulty accessing the specific contents of the actor's reasons, but when they try to present the actor in a positive light, they are motivated to take the actor's perspective and reconstruct the relevant reason contents—what the actor wanted, recognized, or thought about.

Meta-Analysis

In this article we have analyzed multiple hypotheses about actor–observer asymmetries in behavior explanations, three derived from the folk-conceptual model of explanation (the reason hypothesis, belief hypothesis, and marker hypothesis) and two from traditional attribution theory (the person–situation hypothesis and the trait hypothesis). Across multiple studies with variations in methodology and statistical power, we should be able to see clear patterns of support for these hypotheses. No technique integrates empirical data better than meta-analysis, and so we conducted such an analysis on the six studies presented here and three additional actor–observer studies (A1–A3) conducted in our lab, all in all covering data from over 1,300 participants and 8,000 explanations. Details on Studies A1 ($N = 59$) and A2 ($N = 96$) can be found in the supplementary material; Study A3 ($N = 66$) is the control condition reported in Malle, Nelson, Heim, and Knorek (2007). A random-effects model was applied to precision-weighted effect sizes (Hedges & Vevea, 1998; Shadish & Haddock, 1994), using SPSS macros by David B. Wilson, available at http://mason.gmu.edu/~dwilsonb/ma.html. Figure 2 displays the resulting average effect sizes (with 95% confidence intervals [CIs]) for all asymmetries.

The conclusions are clear with respect to the three folk-conceptual asymmetries: All three are reliable, with the reason asymmetry the strongest of the three. For all three, homogeneity

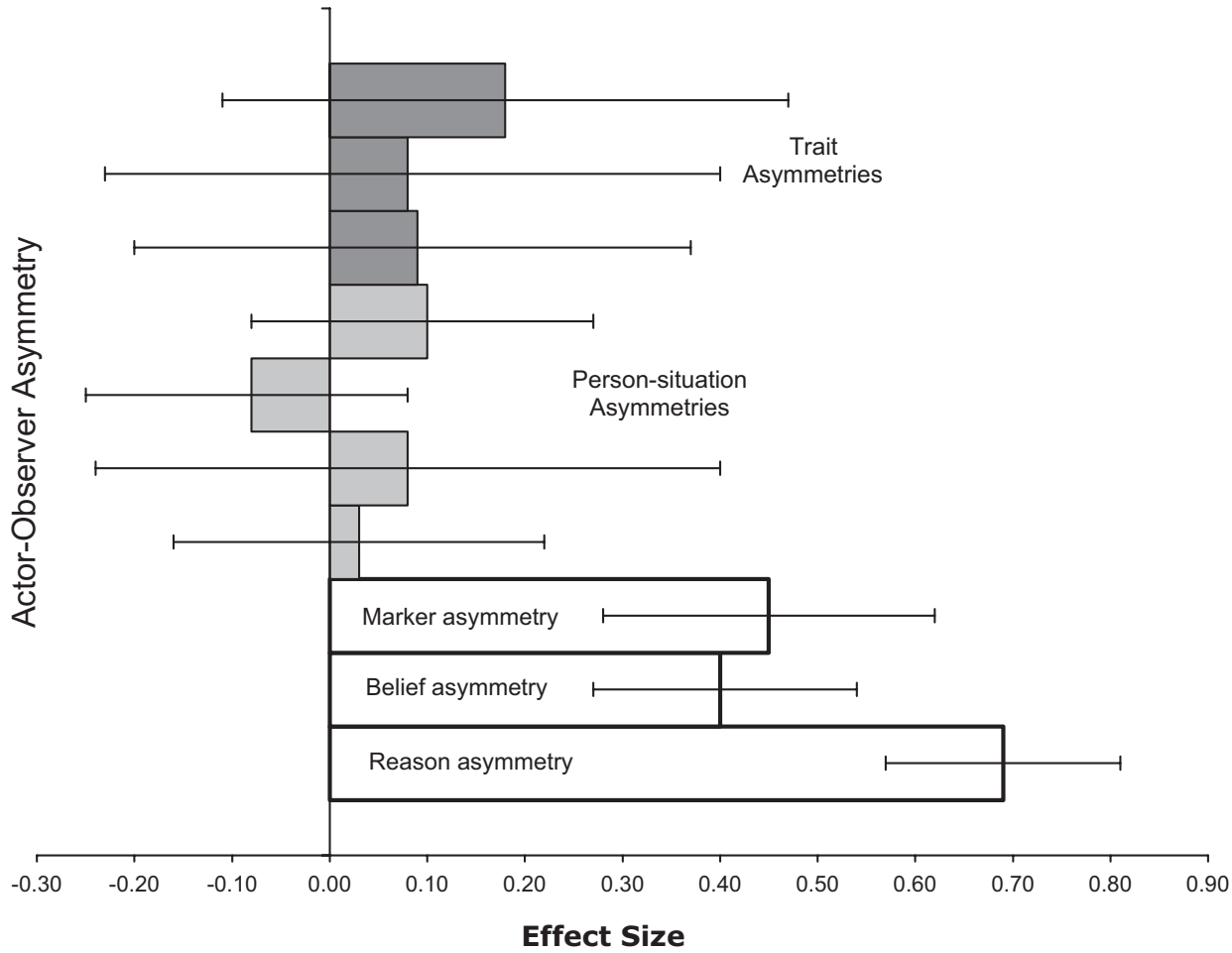

Figure 2. Average effect sizes (and 95% confidence intervals) for all tested actor–observer asymmetries across nine studies.

tests showed no systematic source of variance across studies besides sampling error (Qs = 4.0 to 8.3, ps > .40). We can therefore be quite confident in the estimated true effect sizes: for the reason asymmetry, $\bar{d}$ = 0.69 (95% CI: 0.57, 0.81); for the belief asymmetry, $\bar{d}$ = 0.40 (95% CI: 0.27, 0.54); and for the marker asymmetry, $\bar{d}$ = 0.45 (95% CI: 0.28, 0.62).

Among the traditional hypotheses, there is no compelling evidence for a person–situation asymmetry, either overall or within subtypes of explanations, with average effect sizes ranging from −0.08 to 0.10. The conclusions are also negative with respect to the overall trait asymmetry ($\bar{d}$ = 0.09, *ns*) and its subtypes. These findings are consistent with the meta-analysis of published studies on the traditional attribution hypothesis (Malle, 2006), in which the estimated true effect sizes varied between −0.02 and 0.09.

General Discussion

It would be convenient if there was only one actor–observer asymmetry in behavior explanation. But that is not the case. People's folk explanations of behavior have a complex conceptual structure, comprising multiple modes of explanation and distinct features within each mode. An investigation of actor–observer asymmetries must appreciate this complexity. Accordingly, the present studies examined five hypotheses of actor–observer asymmetries in behavior explanation, displaying solid evidence for three folk-conceptual hypotheses (the reason asymmetry, the belief asymmetry, and the marker asymmetry) but not for either the person–situation hypothesis or the trait hypothesis. We now discuss these findings with a view to the psychological processes that underlie actor–observer differences in behavior explanations.

Processes Underlying Actor–Observer Asymmetries

Many processes have been considered over the years as driving differences between actors' and observers' explanations of behavior, including knowledge, visual perspective, and self-serving motivation. But these processes have not been integrated into a convincing account of the traditional actor–observer asymmetry (Robins et al., 1996), perhaps because the data have been so inconsistent (Malle, 2006) or perhaps because this asymmetry was never well grounded in theory (Buss, 1978; Locke & Pennington, 1982). We have suggested that the person versus situation dichotomy may not be an adequate way to describe what separates actors and observers in the first place, because this distinction neither reflects how people conceptualize human behavior (Buss, 1978; Heider, 1958; Malle, 1999, 2004) nor captures actually existing empirical differences between actors' and observers' behavior explanations (Malle, 2006). We have proposed an alternative theoretical model in which several parameters of explanation characterize the relevant differences in how people explain behavior. Within this model, three reliable actor–observer asymmetries have been demonstrated. The question now becomes what psychological processes underlie these asymmetries.

Reason asymmetry. The choice between reasons and causal history explanations is guided both by processes of cognitive access (what information an explainer can recall, know, or infer) and by the explainer's motivational stance (what effects the explanatory information should have on an audience). We can expect that actors normally have better access to their own reasons than observers do and that they are normally more motivated to portray themselves as active, conscious, and rational agents (which is best done with reasons). But exactly how do these two forces jointly bring about the reason asymmetry, and what constellations of these processes can overcome the asymmetry?

The reason asymmetry has been the strongest across all of our studies (never dipping below d = 0.55), and only in one case did observers offer nearly as many reasons as actors did: when they were explicitly instructed to portray the actor in a positive light. Study 6, because it involved stranger observers, who have no special knowledge, showed that motivation is sufficient to overcome the asymmetry. Two other studies showed that general information access is not sufficient in the same way. When observers were copresent with the actor (Study 4) or generally knew the actor well (Studies 4 and 5), observers' reason explanations did not increase. Future research must therefore establish whether access to more action- and context-specific information (e.g., being privy to the agent's actual deliberations before deciding to act) can overcome the asymmetry.

This asymmetry also touches on a key question in the philosophy of mind: whether actors have privileged access to their own reasons for acting (Gertler, 2003; Wright, Smith, & Macdonald, 1998). If actors offer more reasons merely because of their specific impression-management goals, then the reason asymmetry could be explained without the postulate of privileged access. If, however, actors offer more reasons because they are directly recalling the very reasons they encoded at the time of deliberation (Herrman, 1994), then a complete explanation of the reason asymmetry would have to refer to some form of privileged access, a process fundamentally unavailable to observers (Barresi, 2000; White, 1980). Future research is necessary to distinguish between these two hypotheses.

Belief asymmetry. Once a reason explanation is given, the choice between belief reasons and desire reasons is once more a function of two processes: whether there is cognitive access to the more idiosyncratic information typically represented in beliefs or the more generic information typically featured in desires; and whether the explainer is motivated to portray the agent as rational and thinking or as wanting and needing. Here, too, the default is for actors to have easier access to that idiosyncratic information (Locke & Pennington, 1982) and to be more motivated to use it for impression management purposes. But the data so far suggest that in order to overcome the asymmetry, observers must both gain access to more information and be motivated to use it; neither of the two processes is sufficient on its own. When the information may be available but observers are not necessarily motivated to use it (Study 4 and its follow-up), the belief asymmetry still holds. It also holds when observers are motivated to use such information but are actually lacking the information (Study 6). Only when the requisite information is available *and* observers are motivated to portray the actor in a positive light does the belief asymmetry weaken (Study 5). Further research will have to examine the nature of information that facilitates belief reason explanations. Is it shared appreciation of the context in which the action takes place, or is it access to the specific perceptions and comparisons on which the actor deliberates when deciding to act?

Going beyond actor–observer asymmetries, an intriguing question is how the adult observer's preference for desire reasons over

belief reasons relates to young children's greater ease of using desire reasons rather than belief reasons (Bartsch & Wellman, 1989). Desires derive from goal directedness, which is a conceptual primitive to which infants 6 to 9 months old are sensitive (Woodward, 1998) and that may originate in dedicated neural structures found even in monkeys (Rizzolatti, Fadiga, Fogassi, & Gallese, 1996). In adult cognition, inferring desires or goals from single behaviors appears to be easier and faster than inferring beliefs (Holbrook, 2006), perhaps because goals, more often than beliefs, reveal themselves in bodily motion. To look for the agent's desire or goal may be a fundamental feature of the human social–cognitive system (Gergely & Csibra, 2003; Meltzoff, 1995), and an observer who is asked to explain a behavior may readily represent or search for the action's goal (McClure, 2002). To infer belief, by contrast, an observer will often have to go beyond the motion itself and identify the agent's own representation of the relevant context and options to act.

Marker asymmetry. The current studies identified only one process contributing to the actor–observer asymmetry of using belief markers. We found that (distant) observers who were copresent with actors were as likely to omit belief markers as actors were, whereas absent observers showed a strong asymmetry. For observers, one function of belief markers is to highlight differences in their own and the actor's beliefs; copresent observers share the actor's reality, so there is less need to mark beliefs about this reality. For example,

> (10) "About 15 people came out to help an elderly lady *because the lady was hurt*."

In contrast to copresence, neither intimate knowledge nor impression management attempts curtailed observers' greater use of belief markers. The lack of a knowledge effect was predicted, because knowing more about agents' general considerations should not alter the linguistic phrasing of belief reasons. The lack of an impression management effect was perhaps more surprising. We had previously observed that an explainer's use of a belief marker serves to distance the explainer from the agent's belief (e.g., "He thinks we are getting married"), whereas omission of a marker often indicates an embracing of the agent's belief (e.g., "We are getting married"; see Knobe & Malle, 2002; Malle et al., 2000, Study 6). By extension, one might expect that explainers who try to make the agent look good will more often embrace the agent's belief reasons and therefore omit belief markers. However, a post hoc analysis of belief reason expressions in Study 6 suggested that belief markers can also serve to justify the performed action, and so impression-managing observers may have specifically used marked belief reasons to portray the agent in a positive light. This contention is supported by an extended analysis of belief markers reported in the supplementary material, which also provides evidence for the more general point that observers tend to use belief markers when facing either a psychological distance (disagreement) or a physical distance (noncopresence) from the actor.

Assessing Traditional Hypotheses

The original attribution hypothesis about actor–observer asymmetries was formulated as a contrast between dispositional and situational explanations of behavior (Jones & Nisbett, 1972). The term *disposition* is ambiguous, sometimes referring to stable traits (Jones & Davis, 1965; Shaver, 1975), sometimes to the broader class of "internal causes" of behavior (Kelley, 1967), and this ambiguity leads to two orthogonal hypotheses: Observers may provide more person (relative to situation) explanations than actors do, and within person explanations, observers may provide more traits (relative to nontraits) than actors do.

Our results do not support the person–situation hypothesis. Across the eight studies examined in our meta-analysis, the hypothesized asymmetry was statistically significant only once, hovered around zero in five studies, and reversed twice, resulting in an average effect size of 0.03. This conclusion does not change when we separately consider intentional and unintentional behaviors, for which effect size averages were between −0.08 and 0.10.

Our results also do not support the trait hypothesis, which claims that observers offer more trait explanations than actors do. Nine tests of traits in causal history explanations for intentional behavior showed three asymmetries, three null effects, and three clear reversals. Seven tests of traits in cause explanations showed asymmetries in three cases, two null effects, and two clear reversals. The overall average effect size was 0.09. The range of effect sizes from study to study was substantial ($\bar{d} = -0.60$ to 0.85), which explains why we had speculated in preliminary reports of some of these studies that there may be a trait asymmetry (Knobe & Malle, 2002; Malle, 2002, 2005).

The average effect sizes for the person–situation and trait asymmetries are remarkably similar to those of a recent meta-analysis of 173 published studies on the classic actor–observer asymmetry (Malle, 2006), which averaged between −0.02 and 0.09. Thus, even though textbooks in social psychology have described the classic actor–observer asymmetry as a robust and well-supported phenomenon, there is no evidence for it, either in the published literature or in the present studies.

Two other points are worth mentioning. First, the literature suggested that an actor–observer asymmetry for trait explanations would be weakened when observers are close to and/or knowledgeable about the actor. Our results do not support this hypothesis. In Study 4, there was no trait asymmetry, and knowledge seemed to decrease trait use. In Study 5 there was a reverse trait asymmetry, and knowledge actually increased trait explanations and decreased situational explanations.

Second, not only is there no evidence for a trait asymmetry, people generally use very few traits when explaining behavior (cf. Lewis, 1995; Malle, 2004). Despite people's reputation as "dispositionists" (Ross & Nisbett, 1991), participants in the present studies referred to stable traits in only 5% of all behavior explanations. True, about two thirds of all folk explanations of behavior explicitly referred to "the person," but 80% of these references concerned the actor's mental states. Put differently, 44% of all explanations cited the actor's reasons, and an additional 23% referred to mental states as causes or causal history factors. With two thirds of participants' behavior explanations referring to mental states, but only 5% referring to traits, we must conclude that people are not dispositionists but mentalists. This observation, in contrast to much of social psychological work over the past decades, converges well with developmental, evolutionary, and social neuroscience research, which considers the capacity to represent other people's mental states as the core of social cognition (e.g., Amodio & Frith, 2006; Decety & Grèzes, 2006; Dunbar,

2003; Johnson, 2005; Malle & Hodges, 2005; Mitchell, Macrae, Mason, & Banaji, 2006; Saxe, Carey, & Kanwisher, 2004).

Limitations and Future Directions

The evidence for the three actor–observer asymmetries in folk explanations of behavior can be considered strong and reliable. In addition, we have tried to make some progress toward an understanding of the psychological processes underlying those asymmetries. Impression management, general knowledge, and copresence at the time of action appear to be crucial processes, but each uniquely drives asymmetries, respectively, for reasons, beliefs, and mental state markers. More research is needed, however, on other potential processes, such as conversational rules, valence of the explained behavior, and effects of perspective taking (for observers) and mental state memory (for actors).

One limitation of the present studies is that participants were drawn only from North American culture. Could these same actor–observer asymmetries be found elsewhere? It has been suggested that people from collectivist cultures treat members of their in-group in the same way that people from individualist cultures treat the self (e.g., Al-Zahrani & Kaplowitz, 1993), which might eliminate some actor–observer asymmetries in explanation (Choi & Nisbett, 1998). To further test this claim one might examine behavior explanations that people give for (a) themselves, (b) in-group members, and (c) out-group members. Within individualist communities, we should find all actor–observer asymmetries to hold between explanations of one's own behavior and explanations of in-group or out-group members' behavior. Within collectivist communities, we should find those same asymmetries to hold between explanations of one's own and in-group members' behavior on one side and explanations for out-group members' behavior on the other side. The three distinct folk-conceptual asymmetries permit additional tests that may separate motivational differences from differences in cognitive factors such as thinking styles (cf. Nisbett, Peng, Choi, & Norenzayan, 2001). According to our studies, the reason asymmetry is driven primarily by motivational factors, the marker asymmetry by cognitive factors, and the belief asymmetry by a combination of motivational and cognitive factors. Depending on which asymmetries hold up in cross-cultural comparisons, the evidence would be able to favor either cognitive or motivational accounts of cross-cultural differences in explanation.

The methodology of directly classifying verbal explanations has not been used very often in classic attribution research (but see Fletcher, 1983; McGill, 1989; Orvis, Kelley, & Butler, 1976), and so one might consider it a limitation. However, we believe that most everyday behavior explanations are framed in language because language provides the richest medium to draw the many distinctions that are inherent in people's folk-conceptual framework of behavior. Indeed, the F.Ex coding system, which tries to capture these distinctions, has yielded strong and consistent results in the present as well as previous studies (Malle et al., 2000; O'Laughlin & Malle, 2002). The method is admittedly time consuming, requires training, and demands rigorous coder reliability, but training resources and several data sets are available in the public domain (Malle, 1998/2007). So far the system has been applied successfully in several languages besides English, such as Latvian and Japanese (Dimdins et al., 2005; Teramae & Karasawa, 2007), and it can be used in simplified form for specific research questions (Levi & Haslam, 2005). The strong pairing of theory and method in the folk-conceptual approach also promises to capture data in a wide variety of contexts, many of which would not be amenable to scale-based measurement. Indeed, the folk-conceptual approach has been applied to perceptions of nonhuman agents (Kiesler et al., in press), medical conversations (Knight & Rees, in press), negotiations (Sinaceur, 2007), restorative justice (Nelson, 2003), and intergroup perception and conflict (Dimdins et al., 2005; Teramae & Karasawa, 2007).

Several other potential domains of application of the folk-conceptual theory and method come to mind. As one expression of dehumanization, people may explain others' behavior in the most primitive ways, denying both rationality and intentional agency (Haslam, 2006). The strategic use of reason explanations more generally plays a role in persuasion and propaganda (Malle, 2004, chap. 8). The folk-conceptual approach could also assess subtle changes in explanations as indicators of change in marital therapy and in cognitive treatments of depression, paranoia, and posttraumatic stress disorder. Likewise, autism and schizophrenia have yet to be examined with a view to behavior explanations, which is of particular interest because in these syndromes, theory of mind appears to be compromised. We can expect variations in the conceptual, cognitive, and linguistic aspects of behavior explanations that both express and possibly maintain such illnesses. A standard person–situation model simply cannot make sense of the complexity of these phenomena. Whether the folk-conceptual theory is sufficient remains to be seen; but it does take a serious step toward understanding the complexity.

Several questions posed within the context of traditional attribution research have yet to be examined within the folk-conceptual theory. McClure's (1998) analysis of discounting effects opens the possibility that specific modes or types of explanation might compete with each other and lead to reduced trust in one or another explanation. A potentially helpful distinction offered by the folk-conceptual approach contrasts within-type competition (e.g., one desire reason against another) with between-type competition (e.g., a reason explanation against a causal history explanation). The theoretical underpinnings of discounting, however, are not very precisely formulated (McClure, 1998), and it remains to be seen whether they permit specific predictions that could be mapped onto explanatory parameters.

The phenomenon of self-serving explanations may be reexamined in view of the broader set of parameters identified by the folk-conceptual theory and the more general actor–observer asymmetries we have documented. Self-servingness might be expressed, for example, by modulating the use of reasons versus causal histories or beliefs versus desires or marked versus unmarked beliefs. These distinct explanation parameters allow, as in the case of cross-cultural research, a separate analysis of cognitive and motivational contributions to self-servingness. The results of these studies may also inform a new look at the relationship between explanations and responsibility judgments (Weiner, 1995) and especially the potential power of explanations to alter perceptions of responsibility.

Conclusions

Doing justice to complexity. The actor–observer asymmetry in explanations is typically described as a broad effect involving

person versus situation causes, and the simplicity and elegance of this formulation has surely contributed to its appeal (Watson, 1982). By contrast, the present studies offer a more complex pattern of results, involving multiple parameters of explanation and three distinct actor–observer asymmetries. Though one might regret the lack of simplicity, there is no reason to expect simplicity as a mark of social–cognitive phenomena.

Our theories cannot deny that behavior explanations differ for intentional and unintentional behaviors; that the conceptual structure of intentional behavior generates several different modes of explanation and types within those modes; that language reflects those differentiations in sometimes subtle ways; and that at least two powerful psychological processes operate on behavior explanations: finding meaning in human behavior and managing social interactions (Malle, 2004). Once we apply these pieces to the phenomenon of actor–observer asymmetries, a strong and remarkably consistent picture emerges, counter to the attribution literature, which has not documented a reliable actor–observer asymmetry (Malle, 2006). Asymmetries in fact exist for three parameters of behavior explanation, and each is governed by distinct psychological processes stemming from the broader forces of information access and impression management. To identify these strong and reliable asymmetries, the analysis of free-response explanations has proven highly useful. Staying close, in this way, to people's actual behavior of offering explanations provides maximal flexibility in studying the phenomena of interest in the field as well as in the laboratory.

Carving explanations at their joints. There is an infinite number of ways scientists can divide up classes of explanations. But what we are looking for are the psychologically significant distinctions—the different types of explanations that people select to serve their purposes and that in fact evoke different responses in their audience. By studying actor–observer asymmetries in explanation, we can learn something quite general about these concepts and distinctions that underlie people's explanations of behavior. The folk-conceptual theory proposes that behavior explanations be divided into discrete modes (such as reasons and causal histories) and, within these modes, into specific features (such as belief reasons and mental state markers). The fact that these distinctions have predictive power—here for actor–observer asymmetries, in other studies for group–individual asymmetries (O'Laughlin & Malle, 2002) or rational self-presentation (Malle et al., 2000)—suggests that the folk-conceptual theory of explanation captures the breakpoints, or joints, in the human endeavor of explaining behavior.

References

Abraham, C. (1988). Seeing the connections in lay causal comprehension: A return to Heider. In D. Hilton (Ed.), *Contemporary science and natural explanation: Commonsense conceptions of causality* (pp. 145–174). New York: New York University Press.

Al-Zahrani, S. S., & Kaplowitz, S. A. (1993). Attributional biases in individualistic and collectivistic cultures: A comparison of Americans with Saudis. *Social Psychology Quarterly, 56,* 223–233.

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

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268–277.

Antaki, C. (1994). *Explaining and arguing: The social organization of accounts.* Thousand Oaks, CA: Sage.

Aronson, E. (2002). *The social animal* (8th ed.). New York: Worth.

Baird, J. A., & Baldwin, D. A. (2001). Making sense of human behavior: Action parsing and intentional inference. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 193–206). Cambridge, MA: MIT Press.

Baron, R. A., Byrne, & Branscombe, N. (2006). *Social psychology* (11th ed.). Boston: Pearson.

Barresi, J. (2000). Intentional relations and divergent perspectives in social understanding. *Arob@se: Journal des Lettres et Sciences Humaines, 4,* 74–99.

Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60,* 946–964.

Bernstein, D. A., Clarke-Stewart, A., Roy, E. J., & Wickens, C. D. (1997). *Psychology* (4th ed.). Boston: Houghton Mifflin.

Bruner, J. S. (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.

Call, J., & Tomasello, M. (2005). Social cognition. In D. Maestripieri (Ed.), *Primate psychology* (pp. 234–253). Cambridge, MA: Harvard University Press.

Choi, I., & Nisbett, R. E. (1998). Situational salience and cultural differences in the correspondence bias and actor-observer bias. *Personality and Social Psychology Bulletin, 24,* 949–960.

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

Davis, S. F., & Palladino, J. F. (2004). *Psychology* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Decety, J., & Grèzes, J. (2006). The power of simulation: Imagining one's own and other's behavior. *Brain Research, 1079,* 4–14.

Dimdins, G., Montgomery, H., & Austers, I. (2005). Differentiating explanations of attitude-consistent behavior: The role of perspectives and mode of perspective taking. *Scandinavian Journal of Psychology, 46,* 97–106.

Donellan, K. S. (1967). Reasons and causes. In B. Edwards (Ed.), *Encyclopedia of philosophy* (Vol. 7, pp. 85–88). New York: Macmillan.

Dunbar, R. (2003). The social brain: Mind, language and society in evolutionary perspective. *Annual Review of Anthropology, 32,* 163–181.

Fiske, S. T. (2004). *Social beings: A core motives approach to social psychology.* Hoboken, NJ: Wiley.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Fletcher, G. J. O. (1983). The analysis of verbal explanations for marital separation: Implications for attribution theory. *Journal of Applied Social Psychology, 13,* 245–258.

Franzoi, S. A. (2006). *Social psychology* (4th ed.). Boston: McGraw-Hill.

Galibert, C. (2004). Some preliminary notes on actor–observer anthropology. *International Social Science Journal, 56,* 455–466.

Galper, R. E. (1976). Turning observers into actors: Differential causal attribution as a function of "empathy." *Journal of Research in Personality, 10,* 328–335.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7,* 287–292.

Gertler, B. (2003). *Privileged access: Philosophical accounts of self-knowledge.* Burlington, VT: Ashgate.

Goffman, E. (1959). *The presentation of self in everyday life.* Garden City, NY: Doubleday.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories.* Cambridge, MA: MIT Press.

Gould, R., & Sigall, H. (1977). The effects of empathy and outcome on

attribution: An examination of the divergent-perspective hypothesis. *Journal of Experimental Social Psychology, 13,* 480–491.

Gray, P. O. (2002). *Psychology* (4th ed.). New York: Worth.

Harré, R. (1988). Modes of explanation. In D. J. Hilton (Ed.), *Contemporary science and natural explanation* (pp. 129–144). Brighton, England: Harvester Press.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10,* 252–264.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.

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

Herrman, D. J. (1994). The validity of retrospective reports as a function of the directness of retrieval processes. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 21–37). New York: Springer.

Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 107,* 65–81.

Hirschberg, N. (1978). A correct treatment of traits. In H. London (Ed.), *Personality: A new look at metatheories* (pp. 45–68). New York: Wiley.

Holbrook, J. (2006). *The time course of social perception: Inferences of intentionality, goals, beliefs, and traits from behavior.* Unpublished doctoral dissertation, University of Oregon.

Jin, Z., & Bell, D. A. (2003). An experiment for showing some kind of artificial understanding. *Expert Systems, 20,* 100–107.

Johnson, M. H. (2005). The ontogeny of the social brain. In U. Mayr, E. Awh, & S. W. Keele (Eds.), *Developing individuality in the human brain: A tribute to Michael Posner* (pp. 125–140). Washington DC: American Psychological Association.

Jones, E. E. (1976). How do people perceive the causes of behavior? *American Scientist, 64,* 300–305.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic 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* (pp. 79–94). Morristown, NJ: General Learning Press.

Kashima, Y., McIntyre, A., & Clifford, P. (1998). The category of the mind: Folk psychology of belief, desire, and intention. *Asian Journal of Social Psychology, 1,* 289–313.

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. .

Kenrick, D. T., Neuberg, S., & Cialdini, R. (2006). *Social psychology* (3rd ed.). Boston: Allyn & Bacon.

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.

Kidd, R. F., & Amabile, T. M. (1981). Causal explanations in social interaction: Some dialogues on dialogue. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 3, pp. 307–328). Hillsdale, NJ: Erlbaum.

Kiesler, S., Lee, S., & Kramer, A. (in press). Relationship effects in psychological explanations of nonhuman behavior. *Anthrozoos.*

Knight, L. V., & Rees, C. E. (in press). "Enough is enough, I don't want any audience": Exploring medical students' explanations of consent-related behaviours. *Advances in Health Sciences Education, Theory, and Practice*.

Knobe, J., & Malle, B. F. (2002). Self and other in the explanation of behavior: 30 years later. *Psychologica Belgica, 42,* 113–130.

Kruglanski, A. H. (1975). The endogenous–exogenous partition in attribution theory. *Psychological Review, 82,* 387–406.

Lahey, B. B. (2003). *Psychology* (8th ed.). Boston: McGraw-Hill.

Lalljee, M., & Abelson, R. P. (1983). The organization of explanations. In M. Hewstone (Eds.), *Attribution theory: Social and functional extensions* (pp. 65–80). Oxford, England: Basil Blackwell.

Larsson, P., Västfjäll, D., & Kleiner, M. (2001). The actor–observer effect in virtual reality presentations. *CyberPsychology and Behavior, 4,* 239–246.

Levi, M., & Haslam, N. (2005). Lay explanations of mental disorder: A test of the folk psychiatry model. *Basic and Applied Social Psychology, 27,* 117–125.

Lewis, P. T. (1995). A naturalistic test of two fundamental propositions: Correspondence bias and the actor–observer hypothesis. *Journal of Personality, 63,* 87–111.

Locke, D., & Pennington, D. (1982). Reasons and other causes: Their role in attribution processes. *Journal of Personality and Social Psychology, 42,* 212–223.

Malle, B. F. (1994). *Intentionality and explanation: A study in the folk theory of behavior.* Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Malle, B. F. (1998/2007). *F.Ex: Coding scheme for people's folk explanations of behavior.* University of Oregon. Retrieved May 30, 2007, from http://darkwing.uoregon.edu/~interact/fex.html (Original work published 1998)

Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review, 3,* 21–43.

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: Psychology Press.

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

Malle, B. F. (2005). Self–other asymmetries in behavior explanations: Myth and reality. In M. D. Alicke, D. Dunning, & J. I. Krueger (Eds.), *The self in social perception* (pp. 155–178). New York: Psychology Press.

Malle, B. F. (2006). The actor–observer asymmetry in causal attribution: A (surprising) meta-analysis. *Psychological Bulletin, 132,* 895–919.

Malle, B. F. (2007). Attributions as behavior explanations: Toward a new theory. In D. Chadee & J. Hunter (Eds.), *Current themes and perspectives in social psychology* (pp. 3–26). St. Augustine, Trinidad: SOCS, The University of the West Indies.

Malle, B. F., & Hodges, S. D. (Eds.). (2005). *Other minds: How humans bridge the divide between self and other.* New York: Guilford 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., Knobe, J., O'Laughlin, M. J., 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.

Malle, B. F., Moses, L. J., & Baldwin, D. A. (Eds.). (2001). *Intentions and intentionality: Foundations of social cognition.* Cambridge, MA: MIT Press.

Malle, B. F., Nelson, S. E., Heim, K., & Knorek, J. (2007). *Storms revisited: Visual perspective and the actor–observer asymmetry in attribution.* Unpublished manuscript, University of Oregon.

Marsen, S. (2004). To be an actor or to be an observer? A semiotic typology of narrator roles in written discourse. *Semiotica, 149,* 223–243.

McClure, J. (1984). On necessity and commonsense: A discussion of central axioms in new approaches to lay explanation. *European Journal of Social Psychology, 14,* 123–149.

McClure, J. (1998). Discounting causes of behavior: Are two reasons better than one? *Journal of Personality and Social Psychology, 74,* 7–20.

McClure, J. (2002). Goal-based explanations of actions and outcomes. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 12, pp. 201–235). Chichester, England: Wiley.

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.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31,* 838–850.

Meyers, D. G. (2004). *Psychology* (7th ed.). New York: Worth.

Miller, D. T., & Norman, S. A. (1975). Actor–observer differences in perceptions of effective control. *Journal of Personality and Social Psychology, 31,* 503–515.

Mitchell, J. M., Macrae, C. N., Mason, M. F., & Banaji, M. R. (2006). Thinking about others: The neural substrates of social cognition. In J. Cacioppo (Ed.), *Social neuroscience: People thinking about people* (pp. 63–82). Cambridge, MA: MIT Press.

Monson, T. C., & Snyder, M. (1976). Actors, observers, and the attribution process: Toward a reconceptualization. *Journal of Experimental Social Psychology, 13,* 89–111.

Moore, G. E. (1993). Moore's paradox. In T. Baldwin (Ed.), *G. E. Moore: Selected writings* (pp. 207–212). London: Routledge.

Nelson, S. E. (2003). *Setting the story straight: A study of discrepant accounts of conflict and their convergence.* Unpublished doctoral dissertation, University of Oregon.

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.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108,* 291–310.

O'Laughlin, M., & Malle, B. F. (2002). How people explain actions performed by groups and individuals. *Journal of Personality and Social Psychology, 82,* 33–48.

Orvis, B. R., Kelley, H. H., & Butler, D. (1976). Attributional conflict in young couples. In J. H. Harvey, W. Ickes, & R. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 353–386). Hillsdale, NJ: Erlbaum.

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Peterson, C., Schulman, P., Castellon, C., & Seligman, M. E. P. (1992). The explanatory style scoring manual. In C. P. Smith & J. W. Atkinson (Eds.), *Motivation and personality: Handbook of thematic content analysis* (pp. 376–382). New York: Cambridge University Press.

Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition, 98,* 137–155.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1,* 515–526.

Rathus, S. A. (2004). *Psychology: Concepts and connections* (7th ed.). Belmont, CA: Thomson/Wadsworth.

Raviv, A., Silberstein, O., Raviv, A., & Avi, S. (2002). Young Israelis' reactions to the Rabin assassination: Two perspectives. *Death Studies, 26,* 815–835.

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

Regan, D. T., & Totten, J. (1975). Empathy and attribution: Turning observers into actors. *Journal of Personality and Social Psychology, 32,* 850–856.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3,* 131–141.

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.

Rogoff, E. G., Lee, M., & Suh, D. (2004). "Who done it?" Attributions by entrepreneurs and experts of the factors that cause and impede small business success. *Journal of Small Business Management, 42,* 364–376.

Rosenthal, D. M. (2005). *Consciousness and mind.* Oxford, England: Clarendon Press.

Ross, M., & Fletcher, G. J. O. (1985). Attribution and social perception. In G. Lindsey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 2, pp. 73–114). New York: Random House.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation.* New York: McGraw-Hill.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55,* 87–124.

Scott, M. B., & Lyman, S. M. (1968). Accounts. *American Sociological Review, 33,* 46–62.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind.* Cambridge, England: Cambridge University Press.

Sedikides, C., & Strube, M. J. (1997). Self evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 209–269). San Diego, CA: Academic Press.

Semin, G. R., & Manstead, A. S. R. (1983). *The accountability of conduct: A social psychological analysis.* New York: Academic Press.

Shadish, W. R., & C. K. Haddock. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

Shaver, K. G. (1975). *An introduction to attribution processes.* Cambridge, MA: Winthrop.

Shaver, K. G., Gartner, W. B., Crosby, E., Bakalarova, K., & Gatewood, E. J. (2001). Attributions about entrepreneurship: A process for analyzing reasons for starting a business. *Entrepreneurship: Theory and Practice, 26,* 5–32.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B, 13,* 238–241.

Sinaceur, M. (2007). *Suspending judgment to create value: Suspicion and trust in negotiations.* Unpublished manuscript, INSEAD, Fontainebleau, France.

Storms, M. D. (1973). Videotape and the attribution process: Reversing actors' and observers' points of view. *Journal of Personality and Social Psychology, 27,* 165–175.

Taylor, S. E., & Fiske, S. T. (1975). Point-of-view and perceptions of causality. *Journal of Personality and Social Psychology, 32,* 439–445.

Taylor, S. E., & Koivumaki, J. H. (1976). The perception of self and others: Acquaintanceship, affect, and actor–observer differences. *Journal of Personality and Social Psychology, 33,* 403–408.

Taylor, S. E., Peplau, A., & Sears, R. (2006). *Social psychology* (12th ed.). Upper Saddle River, NJ: Prentice Hall.

Tedeschi, J. T., & Reiss, M. (1981). Verbal strategies as impression management. In C. Antaki (Ed.), *The psychology of ordinary social behaviour* (pp. 271–309). London: Academic Press.

Teramae, S., & Karasawa, M. (2007, January). Effects of entitativity on judgments of intentionality and responsibility. Poster session presented at the annual conference of the Society for Personality and Social Psychology, Memphis, Tennessee.

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.

Weiner, B. (1986). *An attributional theory of motivation and emotion.* New York: Springer.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct.* New York: Guilford Press.

Wellman, H. M. (1990). *The child's theory of mind.* Cambridge, MA: MIT Press.

White, P. A. (1980). Limitations on verbal reports of internal events: A refutation of Nisbett and Wilson and of Bem. *Psychological Review, 87,* 105–112.

White, P. A. (1991). Ambiguity in the internal/external distinction in causal attribution. *Journal of Experimental Social Psychology, 27,* 259–270.

White, P. A. (1993). *Psychological metaphysics.* London: Routledge.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69,* 1–34.

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development, 22,* 145–160.

Wright, C., Smith, B. C., & Macdonald, C. (Eds.). (1998). *Knowing our own minds.* Oxford, England: Oxford University Press.

Appendix

Reliabilities Across All Studies

	Study 1		Study 2		Study 3		Study 4		Study 5		Study 6	
Coding feature	%	κ	%	κ	%	κ	%	κ	%	κ	%	κ
Identify explanations					90							
Codability of phrase	91		89		89		94		87		98	
Actor–observer			98	.95	97	.94			97	.95		
All explanation modes	92	.87	92	.86	89	.80	85	.71	85	.91	93	.89
Reason–CHR	87	.74	.90	.68	88	.64	88	.70	86	.65	91	.79
Belief–desire–valuing	88	.81	97	.95	95	.91	89	.81	95	.91	95	.91
Mental state markers	86	.73	98	.97	95	.88	88	.77	95	.90	100	1.0
Person–situation–interaction	95	.62	91	.83	86	.77	92	.71	83	.72	83	.73
Trait–nontrait	89	.76	95	.79	90	.60	98	.68	92	.76	93	.67

Note. Empty cells indicate that no coding had to be performed. Kappa is not reported when the cell in which both coders agreed on the absence of the classified feature (i.e., identifiable explanation, codability unit) was either very small or missing. CHR = causal history of reasons.

Received November 28, 2006
Revision received April 6, 2007
Accepted April 17, 2007 ■