CHAPTER SIX

Time to Give Up the Dogmas of Attribution: An Alternative Theory of Behavior Explanation

Bertram F. Malle

Contents

Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island, USA

Advances in Experimental Social Psychology, Volume 44
ISSN 0065-2601, DOI: 10.1016/B978-0-12-385522-0.00006-8

Abstract

Attribution research has held a prominent place in social psychology for 50 years, and the dominant theory of attribution has been the same for all this time. Unfortunately, this theory (a version of attribution as covariation detection) cannot account for people's ordinary explanations of behavior. The goal here is to present a theory that can. The theory is grounded in the framework of folk concepts children and adults use to make sense of human behavior, a framework that was already anticipated by Fritz Heider. To introduce the theory, I first map out this folk-conceptual framework, provide evidence for its core elements, and develop the cognitive and social features of behavior explanations, with a focus on the unique properties of intentional action explanation. I then apply the theory to a core attributional phenomenon—actor–observer asymmetries in explanation—and chart two additional applications. In light of these results, I summarize the theoretical and empirical reasons to give up these three dogmas of attribution theory: that behaviors are like all other events, that explaining behavior is choosing between person and situation causes, and that such choices are driven by covariation detection.

1. Preface

I would like to provide a brief personal history to set the frame for this chapter. When I first engaged with the literature on attribution, I was stunned by the clash between the traditional attribution claims about behavior explanation and those put forth by philosophy of action scholars since Aristotle. Moreover, I found it ironic that Fritz Heider—universally hailed as the father of attribution theory—offered a view of attribution that was much closer to the philosophy of action scholars than to standard attribution researchers.

The two views clashed over a pair of straightforward questions: How do people conceptualize human behavior, and how do they explain it? Here are, roughly, their contradictory answers:

> According to attribution research, people see behaviors as events in the world, and they explain them by referring to either situation causes or person causes (especially dispositional person causes).
>
> According to Heider and philosophers of action, people see two classes of behavior in fundamentally different ways: "Unintentional" behaviors are akin to physical events and are explained in simple cause-effect terms; "intentional" behaviors result from an actor's reasoning and forming of an intention, and they are explained by the actor's own reasons.

For unintentional behaviors, the two views largely agreed because, conceptually, unintentional behaviors are comparable to any other event in the world and are also explained in similar cause–effect ways. But the two views differed radically in their account of intentional behavior. Whereas philosophers focused on the unique properties of intentional behavior and their unique form of reason explanations, attribution researchers made no mention of either and focused instead on two kinds of causes assumed to apply to any behavior: person and situation.

No empirical research at the time could decide between these two views. Philosophers (and Heider) relied on intuition; attribution researchers relied on rating scales that were fixed to reflect the assumed person/situation distinction. But how were people *actually explaining behavior*? Should it not be possible to study the phenomenon through its frequent and sophisticated expressions in language? Somewhat naively, I began to collect naturally occurring verbal explanations—in research articles, newspapers, novels, and everyday conversations.

My impression was immediate, strong, and twofold: It was patently obvious that people's explanations of intentional behavior featured what philosophers have called *reasons*: beliefs and desires in light of which the agent decided to act. It was not at all obvious, by contrast, how one would go about classifying people's explanations into "person" and "situation" categories. I encountered ambiguities, felt uncertainty, and quickly discovered well-documented problems of such a classification. But if attribution theory does not apply all that well to natural instances of the phenomenon of "behavior explanation," what can the theory tell us about this phenomenon?

It became clear that a lot of work had to be done. What really are the conceptual assumptions people make about behavior? How do these assumptions guide people's explanations? What antecedents and consequences do these explanations have in social perception and social interaction? In this chapter, I offer the answers to these questions that my colleagues and I have found so far.

2. Introduction

2.1. What attribution?

Attribution research has held a prominent place in social psychology for 50 years. It might be surprising to learn that, as of October 2010, PsycINFO counts 8880 entries with the stem *attributio** in the title since 1960, which is more than *stereotyp** (6447) and *prejudic** (2155) combined. What is even more surprising is the fact that the dominant theoretical framework for attribution research has been the same for 40 years. Such consistency often signals predictive power of the relevant theory; consistent support for its predictions; and the absence of robust theoretical alternatives. In the case of attribution theory, it signals neither. The predictive power of mainstream attribution theories is slight; evidence directly contradicts their fundamental assumptions; and, as I hope to show here, a robust theoretical alternative is available.

If theory replacement is the goal, then we must be clear about what is to be replaced. The problem is that attribution research represents a vast array of work with only loose interconnections. This "attributional supermarket of ideas" (Kruglanski, 1977, p. 592) is difficult to handle as a whole, so to proceed we must observe a basic distinction.

In social psychology, the term *attribution* has at least two meanings. The first, usually labeled *causal attribution*, refers to explanations of behavior—that is, answers to why-questions. The second, typically labeled *dispositional attribution*, concerns inferences of traits from behavior. Even though explanations and trait inferences are related, they are distinct in many ways (Erickson & Krull, 1999; Hilton, Smith, & Kim, 1995; Malle, 2004), and sometimes they are even inconsistent with one another (Johnson, Jemmott, & Pettigrew, 1984). For example, explanations can refer to traits but rarely do (Malle, Knobe, & Nelson, 2007); trait inferences can be explanatory but usually are not (e.g., in typical personality judgment tasks); traits in principle can be inferred from any behavior, whereas explanations are triggered only by surprising or confusing behavior (Malle & Knobe, 1997b).

My focus is on the phenomenon of causal attribution. In this domain, the theory that dominates the textbooks and handbooks is Kelley's (1967) model of attribution as covariation detection, or some variant of it. I will argue that this theory cannot account for people's ordinary explanations of behavior, and my goal is to present a theory that can. To do so, I will return to Heider's foundational steps toward a theory of behavior explanation. Unfortunately, Heider has been so widely misunderstood and so often misrepresented that I will need to reintroduce the reader to Heider's own thoughts about attribution.

Because of my focus on causal attributions, I will set aside Jones and Davis's (1965) correspondent inference theory and the voluminous research literature it spawned (see Gilbert, 1998, for a review). This theory's domain of application is trait inference, not behavior explanation. However, in my later analysis of actor–observer asymmetries in explanation (Section 5.1), I will consider implications that these asymmetries have for trait inference.

2.2. Overview

My plan is as follows. I will delineate what behavior explanations are by returning to Heider's (1958) model of attribution. I will identify the problem he (and attribution research after him) left unsolved and propose a solution. This solution is a theory of behavior explanations that is grounded in the framework of folk concepts that children and adults use to make sense of human behavior. To introduce the theory, I will first map out this folk-conceptual framework, provide evidence for its core elements, and develop the cognitive and social features of behavior explanations. In this I will focus on the previously overlooked unique properties of intentional action explanation. I will then apply the theory to account for a core phenomenon in attribution—actor–observer asymmetries in explanation—and chart two additional applications.

In light of these results, I will summarize the theoretical and empirical reasons to give up the three dogmas of attribution theory, which are: that behaviors are like all other events; that explaining behavior is choosing between person and situation causes; and that such choices are primarily driven by covariation detection.

3. Origins: Getting Heider Right

3.1. Commonsense psychology

Like Lewin and Asch before him, Heider (1944, 1958) recognized that social psychology must chart out people's subjective perceptions of the social world, because these perceptions, whether right or wrong, critically guide social behavior. More than any other social psychologist, Heider focused on the conceptual framework that people rely on when perceiving and explaining human behavior, which he labeled *commonsense psychology*. He took an important first step toward the scientific study of this framework, and decades later, scientists from half a dozen disciplines agree that humans indeed perceive people and their behavior through a unique conceptual framework. This framework, now typically labeled *folk psychology* or *theory of mind*, characterizes behavior as fundamentally linked with mental states. The human recognition that there is mind behind behavior is one of most consequential evolutionary advances, on par with language and probably preceding it (Malle, 2002a; Tomasello, 2003). Heider knew about the importance of folk psychology; and he made substantial strides in describing it, highlighting its complexity, and showing its significance for social interaction.

3.2. Personal causality

But there is a sad fact about Heider's legacy in social psychology. He receives universal credit for "discovering" a much simpler conceptual distinction on which people are said to rely when perceiving and explaining human behavior: the dichotomy between person (or internal) attributions and situation (or external) attributions. Open any social psychology textbook, and this is what students of the discipline learn first and foremost about Heider (e.g., Aronson, Wilson, & Akert, 2010; Kassin, Fein, & Markus, 2008; Myers, 2010). Yet, this dichotomy is clearly not what Heider put forth as the core of commonsense psychology. That core was people's distinction between two very different models of behavior (Heider, 1958, chap. 4): The first is the model of "impersonal causality," which people apply to unintentional human behaviors (e.g., yawning or feeling sad) as well as to physical events (e.g., leaves falling or waves splashing). The second is the model of "personal causality," which people apply to intentional actions. "Personal causality," Heider wrote, "refers to instances in which *p* causes *x* intentionally. That is to say, the action is purposive" (Heider, 1958, p. 100; see also pp. 112, 114). Thus, Heider proposed that the fundamental distinction people bring to social perception is that between *intentional and unintentional behavior*. However, ever since Kelley's (1960) review of Heider's, 1958 book, the word *personal causality* was omitted, replaced by the term "person cause" (or "internal cause"), and its meaning was radically pruned from "intentional" to "any cause inside the person's skin" (Gilbert & Malone, 1995).

One simple move—and decades of attribution research forgot about people's concept of intentionality, arguably a central element of folk psychology and social cognition (Malle, Moses, & Baldwin, 2001; Zelazo, Astington, & Olson, 1999). Contrary to Heider's fundamental insight, people were described as using a plain dichotomy of two causes: For tears, fears, holding hands, and writing letters, "the choice is between external attribution and internal [. . .] attribution" (Kelley, 1967, p. 194). How could such a misunderstanding happen?

And it was a misunderstanding. Nobody ever argued that we should *change* Heider's personal/impersonal distinction to a different distinction; nobody ever offered theoretical reasons or empirical evidence to suggest that people rely on one distinction rather than another. What happened? Maybe this: Listeners are prone to misunderstand when they think they already know what the speaker is going to say. From Kurt Lewin, the field had learned that human behavior is "a function of the person and the situation" (Lewin, 1936, p. 12). That was the scientific viewpoint. But as Kelley and Thibaut (1978) put it: "The man in the street and the scientist share the same general approach to the interpretation of behavior. Both assume that $B = f(P, E)$" (p. 214). Thus, it appeared, Heider must have referred to this dichotomy.

Heider may have contributed to this deep misunderstanding by introducing the intentionality distinction with unusual words ("personal" vs. "impersonal") that still contained the familiar "person–" stem (Malle & Ickes, 2000). Indeed, traces of this confusion of words can be seen to this day. Whereas most scholars falsely claim that Heider himself used the simplified terms "person versus situation," some writers grant him his own original words but prune their meaning: "According to Heider, fundamental to the question of why someone behaves as he or she does is whether the locus of causality for that behavior is within the person (personal causation) or outside the person (impersonal causation)" (Fiske, 2008, p. 140). Not according to Heider.

3.3. Action outcomes versus action explanation

Granted, we can find passages in Heider (1958) in which he referred to a basic person–environment distinction (pp. 56, 82). However, in elaborating on this distinction (pp. 82–87), he made clear that it was not meant to capture the folk theory of *action*. Instead, it applies to the action *outcome,* the result of an action, such as passing a test or reaching the other side of the river. An action outcome is achieved when the agent *tries* to and *can* bring about an outcome by performing an action, but only to the *can* aspects (the outcome-enabling causes) did Heider apply the distinction between person forces (e.g., effort, ability) and environmental forces (e.g., task difficulty, luck). "Whether a person tries to do something and whether he has the requisite abilities to accomplish it are so significantly different in the affairs of everyday life that naive psychology has demarcated those factors" (p. 82). Abilities and other enabling causes (some in the environment) are on the far side of this demarcation; personal causality is on the near side, involving the motivational concepts of intention, desire, goal, and reason (e.g., pp. 100, 114).

In an interview, Heider explicitly distinguished between these two types of attribution tasks (Harvey, Ickes, & Kidd, 1976, p. 14). For one, there is the attribution of *outcomes* (e.g., success and failure), to various causes, which Heider noted was well developed in Weiner's work (e.g., Weiner, 1972). For another, there is the attribution of actions to motives—"answers to the question, 'Why did he do it?', not 'Why did he succeed?'" (p. 14). These why-questions about motivation, Heider maintained, had "not yet been adequately treated in attributional terms."

Figure 6.1 summarizes Heider's model of attribution. The core distinction lies between personal causality and impersonal causality. The impersonal causality area subsumes two subareas: basic cause–effect relations for unintentional events (physical, physiological, or psychological; see Heider, 1958, pp. 146–149) and the process of an action enabling an outcome. The unique involvement of intention occurs only in personal causality, in the generation of intentional action; outcomes that such actions cause in the world are outside the agent's direct control. Neither Heider nor subsequent attribution research clarified in detail the crucial steps that precede intentions: how people explain *why* the agent decided to act, "the reasons behind the intention" (Heider, 1958, p. 110). This chapter tries to supply this long missing theory of action explanation (Lewin, 1936, p. 12).

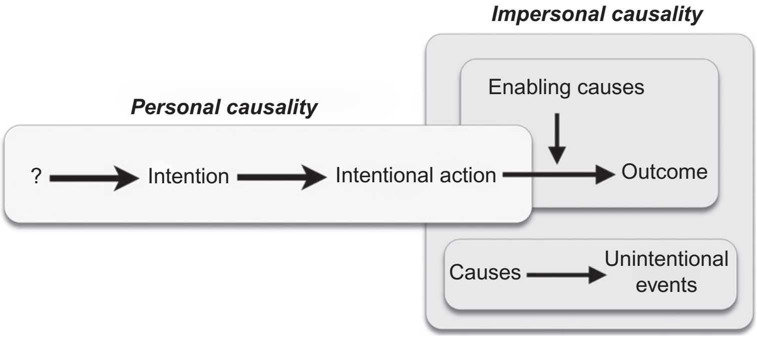

Figure 6.1 Heider's actual model of attribution.

4. A Folk-Conceptual Theory of Action Explanation

Explanations are answers to why-questions. They can be in the head—as private processes of understanding and finding meaning—and they can be in social interaction—as communicative acts of impression management (Hilton, 1990; Malle, 2004). But private and communicated explanations are anchored in the same conceptual framework: the folk concepts people have about behavior. This is why I use the label "folk-conceptual" for the proposed theory of behavior explanation. If we want to know how people explain behavior, we need to know how they conceptualize behavior in the first place (Malle, 1999, 2004; Malle, Knobe, O'Laughlin, Pearce, & Nelson, 2000).

The theory has three layers. The first specifies people's conceptual framework of behavior and derives the explanatory tools that build on this framework—in particular, the distinct modes and types of explanation that fall out of the folk concept of intentionality. The theory's second layer specifies the psychological processes that operate on this framework—processes that select for, generate, and modulate explanations. The third is the linguistic layer, which specifies the tools in language that permit people to communicate explanations and use them for social goals.

I begin with the conceptual layer.

4.1. The folk theory of mind and behavior

4.1.1. General characterization

The central categories of the folk theory of mind are *agent*, *intentionality*, and *mind*, and they are closely related to one another (D'Andrade, 1987; Malle, 2005a). Agents are entities that can act intentionally, intentional actions require a particular involvement of the mind, and only agents have minds. At first glance seeming circular, this web of concepts is anchored by specific perceptual-cognitive processes. For example, objects perceived as self-propelled and behaving contingently are classified into the category *agent* (Johnson, 2000; Premack, 1990). Having identified an agent, the human perceiver is sensitive to face, gaze, and motion patterns that reveal whether the agent's behavior is intentional (Dittrich & Lea, 1994; Phillips, Wellman, & Spelke, 2002), and further analyses of behavior and context lead to inferences of specific goals, beliefs, and emotions (Malle, 2005b).

The intentionality concept is the hub of the folk-conceptual framework. It separates the entire realm of behavior into intentional and unintentional events, guides perceptual and cognitive processes (such as inference and explanation), and influences judgments of praise, blame, and moral responsibility (Cushman, 2008; Guglielmo, Monroe, & Malle, 2009; Malle, 2006a; Shaver, 1985). Many linguists count the concept of intentionality as fundamental to the way humans see the world, and linguistic forms of this concept have been found across all known languages (Bybee, Perkins, & Pagliuca, 1994; Jackendoff & Culicover, 2003). Exactly what do we know about the intentionality concept?

4.1.2. Intentionality: Concept and judgments

Developmental change in the folk-conceptual framework occurs primarily as a differentiation of the intentionality concept from simple categorizations to complex judgments. Infants initially understand intentional behavior as goal-directed—as behavior that is directed toward objects and continues to "aim" at them even when their locations change (Wellman & Phillips, 2001; Woodward, 1998). At the end of the first year, infants are able to pick out intentional actions from streams of behavior (Baldwin, Baird, Saylor, & Clark, 2001). Through the second and third years of life, this behavior-based intentionality concept breaks up into separate mental components, including *desire*, *belief*, and *intention* (Astington, 2001; Wellman & Woolley, 1990; Wimmer & Perner, 1983). But where does it end up? What does a mature understanding of intentionality look like?

In Malle and Knobe (1997a), we offered a systematic study of the adult concept of intentionality. We first showed that social perceivers have a high level of agreement in their intentionality judgments (Malle & Knobe, 1997a, Study 1). Participants read descriptions of 20 behaviors (e.g., "Anne is in a great mood today"; "Anne interrupted her mother") and rated them for their intentionality. About one half of the participants received no definition of intentionality before they made their ratings; the other half did receive such a definition (it means that the person had a reason to do what she did and that she chose to do so). Agreement of intentionality ratings across the 20 behaviors was high: on average, any two people's ratings correlated at $r = 0.64$, and any one person's ratings correlated at $r = 0.80$ with the remaining group, corresponding to a Cronbach α of 0.99. More important, the experimenter-provided definition had absolutely no effect on agreement. These results suggest that people share a folk concept of intentionality and spontaneously use it to judge behaviors. (For a replication in Japan, see Ohtsubo, 2007; in China, see Ames et al., 2001.)

What are the components of the intentionality concept—the features that people require to consider a behavior intentional? On the basis of content coding and experimental data, we developed a five-component model of the folk concept of intentionality, displayed in Fig. 6.2 (Malle & Knobe, 1997a). For a behavior to be judged as intentional, the behavior must be based on a *desire* for an outcome, *beliefs* about the action causing this outcome, a resulting *intention* to perform the action, as well as *skill* and *awareness* when actually performing the behavior.

This multicomponent model was supported in several experiments in which most components were present (either explicitly stated or obviously implied) and one or two were absent. Intentionality judgments were uniformly low when even just one component was absent (0–26%) but high when all components were present (60–97%).

A particularly important feature of the model is that people are said to distinguish between *intention* as a mental state (inferred if beliefs and desires are present) and *intentionality* as a property of an action (inferred when intention as well as skill and awareness are present). In Study 3 of Malle and Knobe (1997a), we presented a story about a group of friends who were debating what movie to go to; they let a coin flip settle the issue, which our protagonist performed. We manipulated the components of skill, belief, and desire, and awareness could easily be inferred. The protagonist was, or was not, able to make a coin land on the side he wanted (skill); knew, or did not know, which side of the coin would favor which movie (belief); and wanted to go to one movie or the other (desire). We then asked participants to make judgments about the protagonist's intention and the intentionality of his action. Table 6.1 shows that for inferring an intention, belief, and desire information is critical; but for judging the action's intentionality, skill information (along with implied awareness) must be present as well.

Little is known about the actual processes underlying intentionality judgments. In subsequent studies, we found that people consider the state of

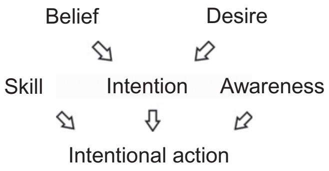

Figure 6.2 The folk concept of intentionality. Reprinted from Malle and Knobe (1997a), with permission from Erlbaum.

Table 6.1 Percentage of Yes responses for trying to get tails and getting tails intentionally with manipulated components of intentionality (reprinted from Malle and Knobe, 1997a, with permission from Erlbaum)

	Judgment	
Components present	Intention (%)	Intentionality (%)
Desire	21	0
Belief	31	0
Desire + belief	81	3
Desire + belief + skill	96	76

Note: The presence of the awareness component could be inferred.

intention as a commitment to act that flows from a reasoning process, and, in this reasoning process, the agent weighs a number of beliefs and desires and settles on a course of action (Malle & Knobe, 2001; see also Mele, 1992). However, we should not expect people to deliberate about the five components of intentionality each time they judge a behavior. In many everyday situations, they will use a more efficient path to assess intentionality. This path involves fast, configural, and often unconscious processing, which may rely on structures developed early in life (Woodward, 2009) and evolved long ago (Call & Tomasello, 1998). Visually perceived behaviors will be considered intentional if they simply "look" intentional (with specific features of biological motion and object-directedness; Baird & Baldwin, 2001; Bertenthal, 1993). Throughout development, humans learn a large number of prototypes of intentional action, and language codifies many of these prototypes in action verbs that imply intentionality (e.g., *reach*, *walk*, *look*, *help*; Malle, 2002b).

When the stimulus behavior is not an immediate prototype match, when critical information is missing, or when perceivers face high demand for confidence and accuracy, judgments will slow down and people will more systematically consider each component of the intentionality concept. This certainly occurs when jurors must make judgments of intentionality in criminal cases (Malle & Nelson, 2003) and for particularly puzzling behaviors in everyday life.

4.1.3. Conceptual foundations of action explanation

The concept of intentionality and its critical mental state components of belief, desire, and intention lay the foundation for people's folk explanations of behavior. We can outline this foundation in four postulates (Malle, 2004),[1] schematically displayed in Fig. 6.3.

[1] As mentioned earlier, I set aside here the treatment of unintentional behavior explanations. The corresponding postulate, however, is: For behavioral events considered unintentional, people use *cause explanations*, which follow the basic cause–effect framework that also applies to physical events.

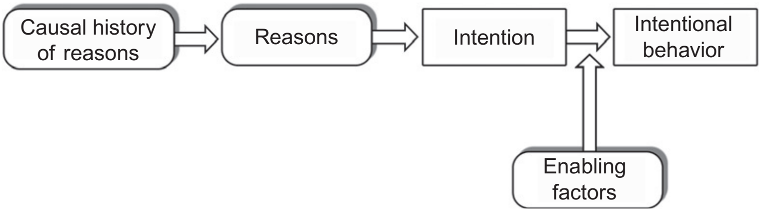

Figure 6.3 Three modes of explanation people use for intentional behavior. Adapted from Malle (1999), with permission from Sage.

1. For behaviors considered intentional, people use one of three modes of explanation: reason explanations, causal history of reason (CHR) explanations, and enabling factor explanations.
2. *Reason explanations* cite an agent's reasons for intending to act or for acting intentionally. *Reasons* are conceptualized as mental states in light of which and on the grounds of which the agent formed an intention to act. For example, "She told me to stay away from the neighbors' kids because *she knew they were a bad influence on me*."
3. *CHR explanations* cite factors that lay in the background of an agent's reasons but were not themselves reasons. Thus, the agent did not form an intention in light of or on the grounds of such causal history factors. For example, "She told me to stay away from the neighbors' kids because *we lived next to this really run-down apartment building* and *she was a bit overprotective*."
4. *Enabling factor explanations* do not clarify why the agent intended to act but rather how it was possible that the intention turned into a successful action. Presupposing that the agent had reasons and an intention to act, they cite factors that enabled the action to be performed as intended. For example, "She worked through the night because *she had a lot of coffee*." I now discuss each of these postulates, and the evidence to support them, in detail.

4.2. A core phenomenon ignored: Reason explanations

Unlike traditional attribution theory, most literatures on behavior explanation have identified reasons as a core phenomenon of explanation (Davidson, 1963; Dretske, 1988; Schank & Abelson, 1977). In particular, the developmental literature on social cognition tracks children's progress from nonverbal recognition of intentional action to verbal explanations of action by referring to specific mental states of beliefs and desires (Bartsch & Wellman, 1995). It is rather unlikely that when children grow up, they forget their mentalistic concepts and suddenly explain behavior using a person–situation dichotomy. Why then have mainstream attribution theories described people as being fixated on person and situation causes?

While this may be puzzling, even more puzzling is the fact that a number of attribution researchers have over time stressed the importance of reason explanations without gaining any measurable impact on mainstream attribution models. Heider stated that people explain why a person is trying to do something by referring to the "reasons behind the intention" (Heider, 1958, p. 110; see also pp. 125–129). Even Jones and Davis (1965) began their seminal article on dispositional inference with a short section entitled "The Naive Explanation of Human Action: Explanation by Attributing Intentions." There the authors attempted to account for "a perceiver's inferences about what an actor was trying to achieve by a particular action" (p. 222), for the process of finding "sufficient reason why the person acted" (p. 220). Despite an apparent plan to present a theory of how actions are explained by reasons, this was the last that Jones and Davis wrote about action explanations. (At least that is true of Jones; Davis (2009) recently discussed the folk-conceptual model I present here.) Likewise, Schneider, Hastorf, and Ellsworth (1979) proposed that explanations for intentional behavior are distinct from those for unintentional behavior; for intentional behavior, they argued, people make "purposive attributions" (p. 127). However, this explanation mode was not examined in any detail.

Around that time, Buss (1978) argued that ordinary people do not explain all behavior with causes (as Kelley had suggested) but rather use reasons to explain intentional behavior. Reasons and causes are fundamentally different types of explanation, Buss argued, and attribution theory confounded the two. Buss's article drew rather negative responses (e.g., Harvey & Tucker, 1979; Kruglanski, 1979), and mainstream attribution theory remained unaffected.

In the following decade, more authors argued that reasons constitute an autonomous form of explanation and that attribution theories must incorporate reasons and goals into their conceptual repertoire (Lalljee & Abelson, 1983; Locke & Pennington, 1982; Read, 1987; White, 1991). Despite these attempts to reform attribution theory, no such reform occurred. Perhaps, the sheer dominance of the mainstream person–situation model and the flow of research using that model created a resistance against change that historians of science know all too well (Kuhn, 1962). But another force stemming against reform was that the critical positions left too many questions unanswered. Why are reasons used to explain intentional behavior and not unintentional behavior? What makes intentional behaviors so special that they require a unique mode of explanation? And what makes reasons so special that they are distinct from causes, rather than a version of causes? In short, what was missing was an actual theory of reason explanations.

4.2.1. What reason explanations are

The nature of reason explanations derives from people's conceptualization of intentional actions. Such actions are seen as generated by intentions, which in turn emerge from the agent's beliefs and desires (D'Andrade, 1987; Kashima, McKintyre, & Clifford, 1998; Malle & Knobe, 1997b, 2001) (see Fig. 6.4, developed from Fig. 6.2). This conception of the intention formation process has been called "practical reasoning" because it resembles the derivation of a conclusion from two premises: the agent's desire for some outcome and beliefs about what action may lead to that outcome (Harman, 1986; Mele, 1992). Conversely, an intentional action can be explained by reference to its desire and relevant beliefs: "Why did you go running?"—"Because I wanted to get in better shape, and . . . I figured that I can do that by going running."

Schematically:

X decided to A [go running].
From: X wanted O [to get in better shape].
X believed that A [going running] leads to O [getting in better shape].

Thus, it is no linguistic accident that these explanations refer to *reasons*, because they are seen as emerging from a *reasoning* process. Moreover, because there is an agent who arrives at an intention through reasoning, it is no accident that people ask "what were *his* reasons?" or find that "*her* reasons make sense," using personal pronouns to indicate that reason explanations refer to beliefs or desires that *this agent* considered. We never ask, "What were his causes for tripping?" or "Her causes make sense." When citing reasons, explainers are not merely referring to some causal factors that influenced the action in question. Rather, they are picking out what they deem the critical steps in the agent's own intention formation process—the process the agent presumably underwent when she decided to act as she did. Offering reason explanations is therefore an act of perspective taking—of adopting the agent's subjective viewpoint and making sense of what favored the action from that particular viewpoint.

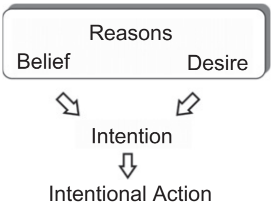

Figure 6.4 Reason explanations derived from the folk concept of intentionality.

The folk-conceptual theory of explanation thus characterizes reason explanations by the following fundamental properties, framed as assumptions that people make about reasons:

When explainers provide *reasons* to explain an agent's behavior, they perceive that behavior to be intentional (intentionality assumption) and try to identify mental states *in light of which* (subjectivity assumption) and *on the grounds of which* (rationality assumption) the agent decided to act.

I discuss evidence for each of these properties in turn.

4.2.2. Reasons apply to intentional behaviors

The first fundamental property is that reason explanations apply only to intentional behaviors and never to unintentional ones. Whenever explainers provide an agent's reasons for acting, the explainers (and their audience, if present) perceive the behavior to be intentional. In support of this proposition, White (1991) demonstrated that people rate behaviors explained by reasons as intentional and behaviors explained by causes as unintentional. However, the different types of explanations (cause vs. reason) in this study were paired with already clearly unintentional versus intentional behaviors. Participants may have made their intentionality judgments based on the behavior, not on the added explanations. In Malle (1999, Study 1), I therefore selected stimulus behaviors whose intentionality was *a priori* ambiguous. Thus, behaviors could be held constant and presented with either a cause explanation or a reason explanation. Any differences in intentionality ratings must then be due to the constraint that the type of explanation (cause vs. reason) puts on the perceiver's interpretation of that behavior.

For this study, I selected two behaviors that had received intentionality ratings around the midpoint in Malle and Knobe (1997a): "Anne was driving above the speed limit" and "Vince interrupted his mother," both of which can be construed as either intentional or unintentional. Each behavior was paired with one of two reason explanations or one of two cause explanations. For example, a reason explanation for Anne's speeding was: ". . .because she knew that the store closed at 6:00," and a cause explanation was: ". . .because she wasn't paying attention to the speedometer." People rated Vince's behavior as reliably more intentional (on a 1–9 rating scale) when explained by a reason ($M = 8.1$) than when explained by a cause ($M = 5.4$), and they also rated Anne's behavior as more intentional when explained by a reason ($M = 7.5$) than when explained by a cause ($M = 4.0$).[2]

[2] The intentionality judgments for cause explanations were only moderately low because the stimulus behaviors, though ambiguous, are not purely unintentional. Even if interrupting is not intentional, the act of saying something is; even if exceeding the speed limit is not intentional, the act of driving is.

4.2.3. Reasons imply subjectivity

The second fundamental property of reasons is *subjectivity,* and it refers to the fact that reason explanations are designed to capture the agent's own reasons for acting (Malle, 1999). Social perceivers normally try to reconstruct the considerations the agent underwent when forming an intention, and they thus take the agent's subjective viewpoint when explaining the action.

Evidence for this property comes from the use of mental state verbs in reasons ("she thought," "he wanted"), in particular when the explainer ascribes to the agent a *false* belief: "Why didn't they send you to a regular boarding school right away?"—"Cause I think at the time they felt that what they were doing was right." Or, "Why did she rush off?"—"She thought she was late for her class." It was the agent's subjective belief, not objective reality, that guided her action and thus explains it.

More direct evidence comes from cases in which a reason explanation is expressed without a mental state verb but people nonetheless assume the agent's subjective awareness of the reason content. Belief reasons are the purest case for such a test because once their mental state verb is omitted, they leave no linguistic trace of being mental states. For example, "She went to Spain for vacation because [. . .] it was hot there," which, according to the subjectivity assumption, is interpreted as "She *thought* it was hot there" (whether or not it really was).

In Malle et al. (2000, Study 1), we tested the tacit subjectivity assumption for explanations like these within an expectation violation paradigm, not unlike that used in developmental psychology. If a perceiver holds a tacit assumption about a stimulus object and if that assumption is violated in an alteration of the stimulus object, then the perceiver should show characteristic responses of rejection or correction. In our study, we examined whether people tended to judge a reason explanation as meaningless if the explanation contradicted the subjectivity assumption—that is, if the agent was said to be unaware of the content of an unmarked belief reason. For example, "Shanna ignored her brother's arguments because they were irrelevant (even though she was not aware that they were irrelevant)." If the reason explanation "they were irrelevant" implies that Shanna subjectively represented her brother's arguments as irrelevant (and for that reason ignored them), then the full statement is meaningless because the parenthetical qualification negates her subjective representation.

We contrasted these reason explanations with CHR explanations whose subjectivity was negated. I will discuss this mode of explanation in Section 4.5, but suffice it to say that CHR explanations do not presuppose an assumption of subjectivity, and people should therefore tolerate negated subjectivity for these explanations. For example, nothing is deeply problematic about the statement, "Anne invited Ben for dinner because she was raised to reciprocate (even though she is not aware that she was raised to reciprocate)."

We presented each participant with four behavior-explanation pairs in which the agent's subjective awareness of the explanation content was negated. Half of the explanations were classified by both expert coders and a group of lay judges as reasons; the other half was classified by expert coders as CHR explanations and by lay judges as "not reasons" (Malle, 1999). With awareness negated, participants judged reason explanations to "make sense" significantly less often ($M = 23\%$) than CHR explanations ($M = 72\%$).

In a follow-up study (Malle et al., 2000, Study 2), we clarified why people still found 23% of reason explanations to be meaningful even when subjective awareness was negated. People read the same behavior-explanation pairs as in Study 1, but this time there was no awareness-negating clause; instead, people were asked a counterfactual prediction question: "If [agent] had not been aware of the fact that *[explanation content]*, would she still have *[action]*?" After saying Yes or No, they clarified their prediction.

In 65% of cases overall, people predicted that, if awareness were absent, an action paired with a reason would not be performed (compared with 31% for actions paired with a CHR explanation). Most important, in 89% of these nonperformance predictions people clarified that it was because the essential reason for performing the behavior was now gone—the agent literally had no reason to act (e.g., "If she didn't know [the plants] needed to be watered, she wouldn't have thought to do it"). In 35% of cases overall, people predicted that, even if awareness were absent, actions paired with reasons would still be performed. But in 98% of these performance predictions people offered alternative explanations for why the action would be performed, because the provided reason was no longer valid (e.g., "A plant still needs water"; "she wants them to be nice looking"). This pattern of results provides strong evidence for the subjectivity assumption as a fundamental property of reason explanations.

4.2.4. Reasons imply rationality

The third fundamental property of reasons is *rationality*, which refers to the fact that the contents of mental states that are cited as reasons have to coherently offer support for the "reasonableness" of the intention and action they brought about. For an intentional action to be adequately explained by reasons, the structure of the above mentioned practical reasoning argument has to be met, which has rationality built in (Davidson, 1982).

Consider an agent who rushes off because "She thought she was late for her class." The action was rationally supported by her belief that she was late for class and by the unmentioned but implied desire to be on time for class. The action would not have been rationally supported if the agent had thought there was plenty of time left or if she had had no desire to be on time. In fact, social perceivers would consider it irrational to rush off in this case, or they would assume that some other reason must sensibly explain the action.

Because we are dealing here with coherence among *contents* of reasons (*what* is desired, *what* is believed), the formal belief–desire–intention structure alone is not sufficient to guarantee rationality. Actual contents have to cohere together in light of people's world knowledge (Lalljee & Abelson, 1983; Schank & Abelson, 1977). In order for a given action explanation to be acceptable, then, the explainer must share knowledge with the agent about how *this action* furthers *this agent's* goals. Conversely, if perceivers face reason explanations that "don't make sense," they should, in line with the rationality assumption, search for additional belief or desire reasons that do support the action. We are currently testing this prediction (Korman & Malle, unpublished data) with stimuli like the following (each of which is shown with one participant's sample response):

She made some chocolate cake because she wanted her brother to develop an appreciation for art.

- Participant's addition: "*It was laid out . . . in different chocolate colors . . . so that it itself was art.*"

She served orange juice to the guests because the roof was leaking.

- Participant's addition: "*She felt bad so she served OJ to the guests to convey some sort of sympathy.*"

4.2.5. Reasons are causal

The characterization of reasons as being subjective and obeying rationality in no way denies that reasons are considered causally generative. As mental states, they are seen as *bringing about* the agent's decision to act. In that sense, people consider reasons to be "causes" (Davidson, 1963; Malle, 1999), but causes with unique properties.

Sometimes reasons are not causal, namely in their role as *potential* reasons when people deliberate about what to do. The agent might consider various goals, weigh their attractiveness, invoke beliefs about how to reach the goals and the consequences of doing so. These are all *possible* but not yet causally effective reasons for acting. Only when the agent in fact formed an intention on the basis of a subset of those beliefs and desires does that subset become the reasons *for which* the agent chose to act. We can speak of *reasons for acting,* without the personal pronoun, when certain reasons speak for one course of action or another ("there are reasons to do that"). We can speak of *the agent's reasons* once the agent decides to act on the basis of the particular set of beliefs and desires—meeting subjectivity and rationality.

4.3. Types of reasons: Desires and beliefs

As we have seen, beliefs and desires are the prototypical reasons because they are necessary elements of the folk concept of intentionality and constitute the agent's practical reasoning toward an intention. But beliefs and desire have distinct conceptual and psychological properties and serve distinct social functions, so we must look at them in more detail.

4.3.1. Desire reasons

Desire reasons reveal the action's preferred outcome, its *goal*, *end*, or *purpose*. Consequently, desire reasons provide the most direct answer to the questions "For what purpose?" and "What for?" By citing a desire reason, the explainer implies that the agent subjectively endorses or values that purpose (Schueler, 2001). For example, "Why did she turn up the volume?"—"To make her brother mad." With this statement the explainer implies that the agent endorses the goal of making her brother mad, so the explainer would be surprised if the agent, in honesty, denied valuing this possible outcome. Also, by citing a desire reason, the explainer points to an as yet unrealized state (of the world or in the agent) and therefore highlights what the agent needs, wants, and hence lacks. This strong agent-focus can be exploited for impression management purposes, such as to make the agent appear selfish (Malle et al., 2000).

Culture and context put considerable constraints on the contents of desire reasons (Bruner, 1990). Even though in principle agents could want anything, in everyday life most will want similar things in a given context. Driving onto the parking lot in front of a grocery store, pulling out one's wallet in front of the cashier, reaching for one's key when coming back to the car—the desire reasons for these actions are easy to infer. That is because they are embedded within schemas and scripts (Schank & Abelson, 1977) and because many desires are revealed in the action itself—through characteristic movements and engagements with objects. We have reviewed earlier that even 6-month-old infants recognize the goal-directedness of such basic actions as reaching (Woodward, 1998), and desires are also the first mental states that toddlers mention in their action explanations (Bartsch & Wellman, 1995).

4.3.2. Belief Reasons

Belief reasons encompass the agent's broad range of knowledge, hunches, and assessments about the action, its outcome, their causal relation, and relevant circumstances. Beliefs are aimed at representing reality and thus are not, by themselves, apt to instigate action. But they are essential in identifying attainable outcomes and selecting appropriate actions with which to pursue those outcomes—they are "the map by which we steer" (Dretske, 1988). Beliefs also help the agent consider the consequences of possible actions, track changes in the desired outcome, and navigate around obstacles. Compared with desire reasons, belief reasons thus highlight the agent's connection with reality, rational deliberation, and states of thinking, rather than the potentially less flattering states of lacking, needing, and wanting.

Compared with desires, beliefs are also more difficult to infer. Culture puts few constraints on the beliefs an agent might have, though context does—copresence with an agent therefore aids the social perceiver in grasping an agent's beliefs. Beliefs, however, do not often reveal themselves in behavior, except for head turns or eye gaze as the specific signs of attention and perception (which are, however, very simple beliefs). Moreover, most actions rely on one goal but many beliefs. Children begin to use belief states in action explanations about a year later than desires (Bartsch & Wellman, 1995), and they build an understanding of beliefs between the ages of 3 and 4, well after they grasp the mental state of desire (Wellman & Bartsch, 1994).[3]

4.4. Linguistic expressions of reasons

A reason explanation can be linguistically expressed in two ways. The explainer may use a mental state verb to mark the type of mental state cited (i.e., a belief or desire) or omit such a verb and directly report the content of the 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: "Because *she wanted* to have an authentic cappuccino." Or he could use the unmarked form: "[____] to have an authentic cappuccino." Likewise, if the explainer chose to cite a belief reason, he could use the marked form: "Because *she thinks* they have the best cappuccino." Or the unmarked form: "Because [____] they have the best cappuccino."

Marked or 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 (Malle, 1999; Malle et al., 2000). This is particularly true for belief reasons.

4.4.1. The significance of marking belief reasons

To say, "I am taking an umbrella because it's raining and. . .," in the same breath, ". . .I don't think it's raining" is logically incoherent (Moore, 1993). Statements of fact ("It's raining") always imply that the speaker believes that fact: "I think it's raining." Therefore, in the first-person (actor) perspective, mental state markers for belief reasons can safely be omitted, as the audience will automatically infer that the unmarked form indicates the speaker's subjective belief.

[3] There is a third, infrequent class of reasons that we have called *valuings*, which encompass liking, disliking, enjoying, etc. (Malle, 1998, 2004, pp. 94–95). A valuing is an affective stance toward a represented event or state, typically directed at something already existing (e.g., "Why did you eat the entire chocolate bar?—I really liked it."). One might be inclined to group them under desires, but our research has so far not shown systematic correlations between valuings and desires, nor have valuings revealed any unique psychological functions.

From the third-person (observer) perspective, however, ascriptions of belief reasons (“She’s taking an umbrella because [she thinks] it’s raining”) operate rather differently. Whether using the unmarked or marked form, the explainer always ascribes to the agent the cited belief (that it’s raining). But the choice between marked or unmarked form gives the explainer the option of declaring whether he *shares* that belief. When using the unmarked form, the explainer declares a fact (“. . . it’s raining”), which implies that he also believes that fact. When using the marked form (“She is packing the umbrella because she thinks it’s raining”), the explainer’s own belief is left unclear, and in some contexts it will imply actual disagreement with the belief. Thus, in the third-person observer perspective, explainers have a choice in their belief reason explanations to endorse the actor’s belief or distance themselves from that belief. Conceptually, the explanation is the same either way (she acted for that reason); pragmatically, however, the observer conveys different social meaning.

We tested this hypothesis in Malle et al. (2000, Study 6). Introducing a pair of men conversing at a party, we had one ask the other. “Jerry, why is your girlfriend refusing dessert?” Jerry then answers in one of two ways: “She thinks she’s been gaining weight” or “She’s been gaining weight.” Participants then indicated how happy they thought Jerry was with his girlfriend’s weight. They easily made the inference that Jerry distanced himself from his girlfriend’s belief in the first (marked) case and was therefore significantly happier with her weight ($M = 5.4$) than in the second (unmarked) case ($M = 2.6$).

This pragmatic significance of mental state markers does not extend to desire reasons. At least in English, the linguistic form of unmarked desires still leaves a trace of the desire state: “Why is she running?”—“So she can be on time” or “To be on time.” Because marked and unmarked desire ascriptions are so similar, they fail to imply anything about the explainer’s endorsement of or distancing from the actor’s desire reason. In fact, a manipulation of mental state markers for a desire reason that Jerry ascribed to his girlfriend (“[She wants] to lose weight”) had no significant impact on people’s inferences (Malle et al., 2000, Study 6).

4.5. Causal history of reason (CHR) explanations

Given that reason explanations are used only for intentional behaviors, are intentional behaviors explained only by reasons? No. Explainers have another option—they can point to factors that lay in the background of those reasons. These factors can be subsumed under the label causal history of reason (CHR) explanations and include such forces as the agent’s unconscious motives, personality, upbringing, culture, and the immediate context (Malle, 1999; Malle et al., 2000; O’Laughlin & Malle, 2002; see also Hirschberg, 1978; Locke & Pennington, 1982).

In the schematic of intentional action explanation, CHR explanations can be located in a layer before reasons (Fig. 6.5). Their explanatory force comes from the narrowing of possible reasons the agent may have had. Consider the following examples: "Anne invited her new neighbor for dinner because she is friendly." "Carey watered the plants because she stayed at home in the morning." Anne's friendliness may have led to a desire to do something nice for her new neighbor or a belief that the neighbor would appreciate it. Carey staying at home in the morning may have led her to realize she had time to water the plants or made her want to do some chores she normally does in the evening. Thus, CHR explanations do not deny that the agent had reasons to act; they just do not directly refer to those reasons. Instead, they refer to causal factors that presumably *led up to* the agent's reasons.

But these CHR factors that led up to the agent's reasons are not themselves reasons. As a result, the crucial assumptions of subjectivity or rationality that define reasons do not apply to causal history factors. The explainer who proposed that "Anne invited her new neighbor for dinner because she is friendly" did not imply that Anne thought to herself: "I am friendly; therefore I should invite her to dinner."

Two pieces of evidence support the contention that causal history explanations presuppose neither subjectivity nor rationality. First, whereas people do not tolerate the agent's lack of awareness of the contents of her reasons, they are very tolerant of the agent's lack of awareness of CHR factors (Malle et al., 2000; see Section 4.2.3). Second, in Malle (1999, Study 3), people were asked to identify reasons in a set of 24 explanations that had been *a priori* categorized as 12 CHR explanations and 12 reason explanations. Reasons were defined as "conscious, deliberate reasons for acting that way" and "something that the agent had on his or her mind when deciding to act," thus highlighting the subjectivity and rationality assumptions. Participants separated the 12 reason explanations from the 12 CHR explanations with nonoverlapping classification rates. In a signal detection analysis, this discrimination amounted to $d' = 2.5$.

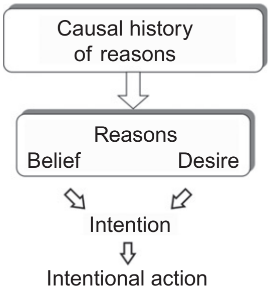

Figure 6.5 Causal history of reason explanations and reason explanations as the two major modes people use to explain intentional actions.

The fact that CHR factors do not presuppose subjectivity and rationality has three corollaries. First, CHR explanations do not rely on an act of perspective taking. Therefore, any observer who cannot or does not want to take the agent's subjective perspective to offer a reason explanation can still offer a CHR explanation (e.g., "Why were they joking at a bum's expense?"—"I cannot explain why they did it other than they were intoxicated"). Second, CHR explanations divert attention away from the agent's reasoning and choice capacity and toward the broader causal nexus in which the action is embedded. Without denying that the action in question was intentional, CHR explanations portray the agent as less rational than reason explanations do (e.g., "When the medical people came by she was telling them what to do *cause she was in panic*"). Third, the permissible disconnection of CHR factors from the agent's subjective perspective allows explainers to cite unconscious motives that the agent would not readily endorse. For example: "Why did he tell you all those intimate things?"—"He was looking for some sympathy." CHR explanations therefore make the agent look less in control than reason explanations do.[4]

Even though CHR explanations and reason explanations are clearly distinct, they are not always in competition with each other—in just over 20% of intentional action explanations they actually co-occur. In these cases, the CHR explanation often provides the necessary background for making the cited reason intelligible. For example, "I decided to go back to school *because when I was in Tahoe they had all these casinos and all these bums wandering around* [CHR]*, and I thought, 'I don't want to be like that'* [reason]."

Enabling factor explanations. The third, and relatively rare, mode of explaining intentional action refers to factors that enabled the action to come about as it was intended. These enabling factor explanations refer to the agent's skill, effort, opportunities, facilitating circumstances, and the like. Whereas reason explanations and CHR explanations focus on clarifying what *motivated* the agent's intention and action, enabling factor explanations take it as a given that the agent had motives and attempt to clarify *how it was possible* that the action was successfully performed. For example, "She hit her free throws because she had practiced all week." This is not a motivational account of why the agent decided to hit the free throws; rather, the agent's practicing all week is identified as the critical factor that allowed her to perform the action she had intended. Consequently,

[4] It is worth pointing out that sociological and psychoanalytic explanations of behavior focus heavily on causal history factors, such as when suicide is explained with reference to social distance or a heavy self-imposed work load is explained with reference to a need for power. Similarly, the increasing emphasis in psychology on priming and unnoticed external stimuli as antecedents even of intentional behavior reflects the same tendency: diverting attention away from the conscious and rational antecedents of actions and portraying people more as hapless nodes in a causal network. In many cases, scholars overlook the fact that they merely describe factors in the causal history of reasons, whereas reasons are still necessary to form an intention and get an intentional action off the ground. For a discussion of these issues, see Malle (2006b).

enabling factor explanations are offered primarily for difficult actions or when the explanatory question is "How was this possible?" rather than "What for?" (Malle et al., 2000; McClure & Hilton, 1997). When the explained event is not the action itself but the outcome it aimed to achieve, then we return to Heider's *outcome attributions*. Both enabling factors and outcome attributions may be categorized according to dimensions such as locus or stability; and Weiner (1986) has shown that these dimensions can be predictively useful, at least in the achievement domain.

4.6. The processes of explanatory choice

We have seen that people have a variety of tools available when trying to explain intentional behavior. But what psychological processes determine people's choices among these explanatory tools? When do people offer CHR explanations instead of or in addition to reason explanations? When do they provide belief reasons, when desire reasons? And under what conditions do they mark their reasons or prefer to leave them unmarked? Trying to answer these questions is the goal of the second layer, the process layer, of the folk-conceptual theory of explanation.

Addressing the questions of explanatory choice requires attention to properties of the various explanatory tools (e.g., of reason explanations, belief reasons) and properties of the explainer (Fig. 6.6). These properties make up two types of processes, reflecting both the cognitive and social nature of explanations (Malle, 2004): There are *information processes*, which govern acts of *devising and assembling* explanations in the mind, and *impression*

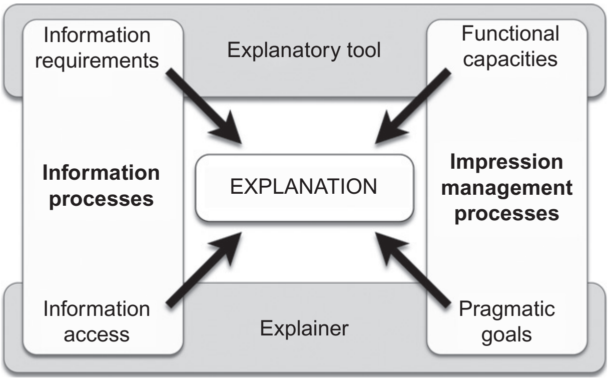

Figure 6.6 Determinants of explanatory choice: information processes (consisting of information requirements of the explanatory tool and the explainer's information access) and impression management processes (consisting of functional capacities of the explanatory tool and the explainer's pragmatic goals).

management processes, which govern the actual *use* of explanations in social interaction.

On the side of information processes, each explanatory tool poses *information requirements* that dictate what information it takes to assemble the particular tool (e.g., a reason explanation); in addition, the explainer has varying amounts of *information access* to assemble the given explanation. Access includes the availability of knowledge about the agent, action, or context; perceptual awareness (e.g., of the agent's moving body); and formats of representation (e.g., the actor is directly aware of the contents of her belief reasons; Rosenthal, 2005).

On the side of impression management processes, each explanatory tool has *functional capacities* that constrain what can be accomplished with the particular tool; in addition, the explainer has various *pragmatic goals*, which define what it is that needs to get accomplished with the given explanation. Such goals include making the agent look socially acceptable, deflecting blame, and either endorsing the agent's reason or distancing oneself from it. In addition, explainers often need to design explanations in light of the audience's knowledge and interest (Kashima et al., 1998; Slugoski, Lalljee, Lamb, & Ginsburg, 1993), which can favor one explanation mode over another (e.g., reasons vs. enabling factors; Malle et al., 2000).

With the process layer of the folk-conceptual theory specified, we can now apply the theory to a variety of phenomena. I will focus on the actor–observer asymmetry, which has been the poster child of attribution theory, and I will discuss two other applications in less detail.

5. Applying the Theory

5.1. Actor–observer asymmetries

Few psychological claims are as convincing as the one that representations of oneself (the "actor perspective") differ in important ways from representations of others (the "observer perspective"). Many self–other differences have been documented, such as in attention (Malle & Pearce, 2001; Sheldon & Johnson, 1993), memory (Rogers, Kuiper, & Kirker, 1977), and personality description (Locke, 2002; Sande, Goethals, & Radloff, 1988). But no difference is better known than the actor–observer asymmetry in attribution, as initially proposed by Jones and Nisbett (1972): "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).

In this section, I will first examine the evidence for the classic actor–observer asymmetry—in particular, the results of a recent meta-analysis that aggregated almost 200 tests of this hypothesis. Then I will develop three alternative hypotheses about potential actor–observer asymmetries within the folk-conceptual theory of explanation and report a series of studies that have confirmed these asymmetries but disconfirmed the classic hypothesis.

5.1.1. The classic actor–observer asymmetry: A robust illusion

The actor–observer asymmetry as proposed by Jones and Nisbett has been described as "robust" (Jones & McGillis, 1976, p. 304), "pervasive" (Aronson, 2002, p. 168), and "firmly established" (Watson, 1982, p. 698). With over 1500 references to the original Jones and Nisbett paper, there can be little doubt that the actor–observer asymmetry in attribution is "an entrenched part of scientific psychology" (Robins, Spranca, & Mendelsohn, 1996, p. 376).

One would expect that such strong claims about the actor–observer asymmetry are backed by equally strong evidence. Surprisingly, no systematic review of this evidence had been conducted since Watson (1982). That article, moreover, covered only a small portion of studies already published at the time, and a large number of additional studies have since become available. I therefore conducted a meta-analysis on the primary literature of published actor–observer studies between 1971 and 2004 (Malle, 2006c).

Across 173 studies in 113 published articles, the classic actor–observer asymmetry yielded average effect sizes between $\bar{d} = -0.015$ and 0.095, depending on statistical models and specific attribution scores (Malle, 2006c). Moreover, the data showed a publication bias. The effect size values were not symmetrically distributed around the mean; extreme negative (hypothesis-disconfirming) values were underrepresented relative to extreme positive (hypothesis-confirming) values. Applying the bias correction by Duval and Tweedie (2000), the average effect size slipped to 0. It seems reasonable to conclude that the classic actor–observer hypothesis is neither robust nor pervasive nor firmly established; the hypothesis appears to be a widely held yet false belief.

But because the classic actor–observer hypothesis was formulated in terms of person and situation attributions, its disconfirmation should not be mistaken for the disconfirmation of actor–observer asymmetries in general. Asymmetries may well exist but cannot be detected by a model of attribution that relies on a framework of person/situation causes.

The folk-conceptual theory of explanation can detect such asymmetries. Below I develop three actor–observer asymmetries that fall out of this theory. In a series of studies, we tested the stability of these asymmetries and compared them with the traditional person–situation hypothesis (Malle et al., 2007). If the traditional predictions are incorrect and the predictions of the folk-conceptual theory are correct, then we have strong arguments both for the existence of actor–observer asymmetries and for the theoretical value of the folk-conceptual approach.

5.1.2. Real actor–observer asymmetries

In Section 4, I described the major tools people use when explaining intentional behavior. These tools demand choices primarily between reason explanations and CHR explanations between belief reasons and desires reasons; and between the use or omission of mental state markers for those reasons. For each of these choices, we can identify two major psychological processes that may differ between actors and observers (see Fig. 6.6): *Information processes* resulting from the interplay between requirements of the explanatory tool and the explainer's information access; *impression management processes* resulting from the interplay between functional capacities of the explanatory tool and the explainer's pragmatic goals. Three actor–observer asymmetries follow from the impact of these processes (Malle et al., 2007).

Hypothesis 1 (Reason Asymmetry)

The informational requirements of reason explanations are these: A reason must explain the specific agent's specific action by obeying the constraints of subjectivity and rationality. For actors, this is typically a straightforward task of recalling the very reasons for which they decided to act (Herrman, 1994). For observers, information access is more challenging; they must rely on stored knowledge, inferences, and simulations, with the real possibility that no explanation can be found. Compared with actors, then, observers will be less often able to provide reason explanations.

The impression management processes push in the same direction. The pragmatic capacities of reason explanations allow the explainer to highlight agency and rationality (Kiesler, Lee, & Kramer, 2006; Malle et al., 2000); because actors are normally more invested in portraying themselves in these favorable ways, actors' use of reasons should exceed that of observers.

Hypothesis 2 (Belief asymmetry)

For actors, the informational requirements for belief reasons will not differ from those for desire reasons because there should be no tendency to directly recall one type of reason better than another. For observers, however, inferring belief reasons will be more difficult, because beliefs require access to more idiosyncratic, less observable information that does not readily derive from social rules and cultural practices (Bruner, 1990). Desires, by contrast, are more bound to such rules and practices and are more immediately visible in human movement (Baird & Baldwin, 2001; Malle, 2005b).

Impression management processes will again push in the same direction. Belief reasons have the capacity to portray the actor as rational and as tracking reality, characteristics that actors would normally like to portray (Malle et al., 2000).

Hypothesis 3 (Belief marker asymmetry)

Belief markers have no information requirements because the same information is expressed here in two different ways—with or without a mental state verb. However, there is a unique access difference between actors and observers. Actors directly represent the *content* of their beliefs (e.g., *the plants are dry*); they do not normally represent their own belief *qua* mental state (they do not represent *I believe the plants are dry*) (Rosenthal, 2005). As a result, when formulating their belief reasons in language, actors will focus on *what* they represented and not *that* they represented something, so they will typically 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, which invites marking those beliefs with a mental state verb: "She thought the plants were dry." A predictable exception to this pattern should be when observers are copresent and fully share the actor's environment, hence the contents of the actor's beliefs. Sitting in a restaurant with my wife and choosing from the same menu, I can explain her choice of taking the duck by directly referring to a desirable (and mutually known) aspect of the chosen option ("It's a creative preparation").

Impression management processes are unlikely to add to the belief marker asymmetry. Even though the choice between a marked or unmarked belief gives the explainer the option of declaring whether he shares the agent's belief, this option is open to both actors and observers. Even actors may sometimes want to distance themselves from their own false beliefs: "Why did you not invite me to the party?"—"Oh, I thought you were out of town!"

Briefly, I should point out that the folk-conceptual theory does not predict an actor–observer asymmetry for cause explanations of unintentional behaviors (either on the basis of a person–situation classification or any other classification). That is because the explanatory tools of cause explanations neither have any specific information requirements nor functional capacities that would favor actors over observers in any way (but see Section 5.2 for their possible role in self-serving explanations). Indeed, both a meta-analysis (Malle, 2006c) and follow-up tests (Malle et al., 2007) show no evidence that actors and observers differ in explaining causes of unintentional behaviors.

5.1.3. Methodology and results

We tested the three actor–observer hypotheses across several eliciting conditions (Malle et al., 2007). In some studies, we asked people to recall previously puzzling behaviors along with their explanations; in others, we extracted spontaneous explanations from people's conversations. In some studies, people provided explanations in private, in others, they communicated them to an audience. In all studies, we classified people's own written or spoken explanations into the explanation types relevant to the hypotheses. To test the traditional actor–observer asymmetry, we also classified all explanations into the categories of person versus situation attributions and, among person attributions, into traits versus nontraits. All in all, six studies provided consistent and strong support for the folk-conceptual hypotheses (with average effect sizes of $d = 0.50–0.70$; see Fig. 6.7) and no support for the traditional hypothesis (with *d*s < 0.12).

5.1.4. Testing specific processes

Three of the studies in Malle et al. (2007) examined some of the hypothesized processes responsible for the asymmetries: variations in information access and impression management goals.

Reason asymmetry. The information access hypothesis suggests that actors can often recall the reasons for their actions whereas observers must rely on stored knowledge, inferences, and simulations. Two studies in Malle et al. (2007) attempted to increase information access for observers and thereby diminish the reason asymmetry. In Study 4, observers explained actions for which they were copresent with the actor when the action occurred (which should aid evidence-based inference); in Study 5, observers explained actions by actors whom they generally knew well (which should aid knowledge-based inference). In neither case did observers' reason explanations increase. Perhaps the information manipulations were not reason-specific and affected CHR explanations just as much; or perhaps the information access difference lies deeper, namely, in the observer's failure

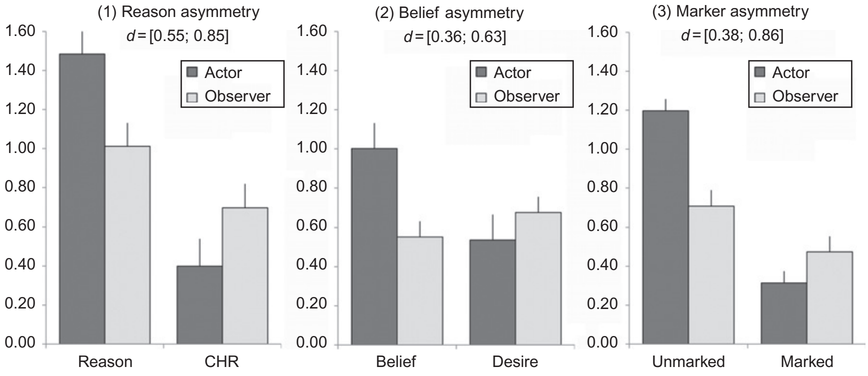

Figure 6.7 Three actor–observer asymmetries predicted by the folk-conceptual theory of explanation with means and 95%-confidence intervals of effect sizes across six studies. The *y* axis refers to the following dependent variables: In (1), to reasons or causal history of reason (CHR) explanations per intentional behavior explained; in (2), to belief reasons or desire reasons per intentional behavior explained by reasons; and in (3), to marked or unmarked beliefs per intentional behavior explained by belief reasons.

to adequately simulate the actor's mental states. Future studies must find ways to manipulate observers' selective access to reasons—for example, by making them privy to the agent's actual deliberations before deciding to act or by specifically instructing them to simulate the actor's perspective. In addition, we might attempt to *decrease* actors' information access—which may occur for actions that they performed long ago and whose reasons may not be directly recalled.

The impression management hypothesis suggests that whereas actors are normally invested in portraying themselves as autonomous, reasoning agents, observers are not as concerned with this task. In Study 6, we changed the observers' motivation and specifically instructed them to explain an agent's behavior and make the person look good in front of an audience. Compared with a control condition without such instructions, where we found the typical reason asymmetry ($d = 1.03$), the impression management condition showed a greatly reduced asymmetry ($d = 0.20$). Because the instructions provided no knowledge gain for the observer, the goal to portray the actor in a positive light appeared to be the sole force in dampening the asymmetry. Exactly how this goal affects explanations is not yet known. Making the actor look good may encourage observers to try harder in taking the actor's subjective perspective—which then reduces the asymmetry by changing information access. In future work, we plan to conduct a fine-grained analysis of the observer's cognitive processes in arriving at a greater number of reason explanations—through perspective taking, simulation, or tailored knowledge retrieval.

Belief-desire asymmetry. Three studies in Malle et al. (2007) spoke to the driving forces of the belief asymmetry. The results suggested that in order to overcome the asymmetry, observers must both gain access to more information and be motivated to use the information for impression management purposes (Study 5); neither of the two processes by itself was sufficient to alter the belief–desire asymmetry (Studies 4 and 6). Further research will have to examine the nature of information that facilitates belief reason explanations. Is it shared appreciation of the objective 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?

Belief marker asymmetry. Intimate knowledge did not curtail observers' greater use of belief markers (Study 5), which was expected because knowing more about agents should not alter the linguistic phrasing of their belief reasons. Somewhat surprising was that instructions to portray the agent in a positive light doubled the strength of the marker asymmetry ($d = 1.18$). A follow-up analysis suggested that observers increased marked beliefs as a way to justify the agent's behavior. For example, "Why did he ask his friends to take him to the hospital?"—"He had good reason to believe he had a serious injury"; "Why did she befriend her?"—"... she realized that the way that she had treated the girl before was wrong." Thus, by choosing specific mental state markers, explainers can either distance themselves from the agent's subjective beliefs (e.g., she thought) or highlight the agent's subjective but accurate connection to reality (e.g., she realized).

Copresence, predicted to curtail the marker asymmetry, indeed showed its impact: Observers who were copresent when the action occurred were as likely to omit belief markers as actors themselves were, whereas observers who were absent when the action occurred showed the usual asymmetry (Study 4). Copresent observers are likely to share the actor's reality, so there is less need to mark beliefs about this reality. For example: "About 15 people came out to help an elderly lady *because the lady was hurt*" (unmarked belief). Everybody saw that the lady was hurt—including the copresent explainer. By contrast, to clarify why "A random nice boy was taking my drunk roommate in and giving up his room for her," the explainer who was not present at the time of the action indicates, "*He realized that she needed help*" (marked belief).

5.1.5. Dispositionism and the fundamental attribution error

By now, the reader may have wondered whatever happened to dispositions—a central term in the history of attribution research. Unfortunately, the meaning of this term shifted from a broad meaning in Heider's work to a narrow meaning in subsequent research. According to Kelley (1960), in his review of Heider's (1958) book, "Heider's central theme is that perception leaps over the raw data presented and enables the person to understand the stable, dispositional properties . . . that account for them" (p. 2). Within Heider's theory, *dispositions* were indeed invariances that perceivers search for when observing behavior. But he considered "motives, intentions, sentiments . . . the core processes which manifest themselves in overt behavior" (p. 34). "A description of movements in terms of motives . . . taps environmental layers of greater invariancy" (Heider & Simmel, 1944, p. 256). By invariance Heider did not mean *stable* or *enduring*. Invariance was a relative term. Relative to the highly variable stream of observed movements, mental states "give meaning to what [the observer] experiences" (p. 81).

The subsequent literature ceased to consider dispositions as mental states. Jones and Davis's (1965) influential theory of correspondent inference fixed the meaning of the term *disposition* to refer to enduring character traits and attitudes. Research on trait inference led to the postulate of the fundamental attribution error (FAE; Ross, 1977); spawned cognitive models of the trait inference process (e.g., Gilbert, Pelham, & Krull, 1988; Quattrone, 1982; Trope, 1986); and examined the possibility of spontaneous trait inferences (Winter & Uleman, 1984; for a review see Uleman, Saribay, & Gonzalez, 2008). What place do such trait inferences have in the folk-conceptual theory of explanation, and what does the nonexistence of the traditional actor–observer asymmetry tell us about the FAE?

Whether the actor–observer asymmetry has implications for the fundamental attribution error depends on exactly what is meant by that error.

If, as many textbooks claim, the FAE is "part of the actor–observer bias" (e.g., Morris & Maisto, 2006, p. 450; see also Deaux & Wrightsman, 1988; Hockenbury & Hockenbury, 2006), then the FAE does not exist, because actors and observers do not differ in their person attributions, situation attributions, or trait attributions (Malle, 2006c; Malle et al., 2007). Similarly, if the FAE is formulated in terms of explanation tendencies within the observer perspective (e.g., "people are inclined to offer dispositional explanations for behavior instead of situational ones," Ross & Nisbett, 1991, p. 125), then the FAE does not exist. That is because people spontaneously refer to stable traits in only 5–10% of all behavior explanations, whereas they refer to the situation in about 20% of all behavior explanations (Malle, 2004; Malle et al., 2007). And as Heider had anticipated, people predominantly offer explanations that refer to *mental states* (in the form of reasons, causes, or causal histories), and these make up 68% of all behavior explanations. We might therefore say that people are not dispositionists, as has been claimed (Ross & Nisbett, 1991); they are mentalists.

However, the FAE need not be equated with the actor–observer asymmetry. When we heed the distinction between trait inference and behavior explanation (see Section 2.1), then the status of the FAE is a separate empirical question. Ross and Nisbett (1991) claim that "People infer dispositions from behavior that is manifestly situationally produced" (p. 126), and people "assume a person has traits corresponding directly to the type of behavior that was exhibited" (p. 88). Defined this way, the FAE can occur when a perceiver observes a behavior that may look diagnostic of an underlying trait (e.g., personality, attitude) but that was, in fact, strongly pressured or enticed by the situation. On this strict interpretation of the FAE as *incorrect trait inferences from single behaviors*, the FAE is not refuted by the absence of actor–observer asymmetries in disposition–situation *explanations* of behavior. It may be disconcerting that observers' explanations so rarely refer to traits, but we simply do not know the prevalence of the FAE—we do not know how many behaviors really are more strongly influenced by the situation than observers believe. Nearly all data on the FAE have emerged from tightly constrained lab experiments, not surveys, observations, or archival studies, and where a naturalistic approach was taken, the results did not look supportive (Block & Funder, 1986; Lewis, 1995). In light of the surprising fate of the classic actor–observer hypothesis, it seems essential to take a fresh and critical look at the prevalence of (erroneous) trait inferences in everyday social contexts.

Beyond actor–observer asymmetries, I now look at two other domains of inquiry to which the folk-conceptual theory of explanation can be applied—one familiar to attribution theorists (the self-serving bias), the other less so: the pattern of explanations people show for the behaviors performed by group agents as opposed to individual agents.

5.2. Shades of self-servingness

The self-serving bias in attribution is usually defined as an actor's reversal of person versus situation attributions for negative compared with positive events. When explaining positive events (e.g., success), actors are predicted to use more person causes, but when explaining negative events (e.g., failures) they are predicted to use more situation causes. The assumed process is that people try to "take credit" for positive events and "deflect blame" for negative events.

The meta-analysis on the traditional actor–observer asymmetry allowed a test for this data pattern by assessing actor–observer differences for negative and positive events, comparing actors against the comparison standard of observers, because the latter are presumably not biased by credit and blame motives. Indeed, the data suggested that, for positive events, actors offered more person causes (and fewer situation causes) than observers did, but for negative events actors offered fewer person causes (though no more situation causes) than observers did.

A number of questions arise, however, about the interpretation of this data pattern. One is what "person attributions" and "situation attributions" actually mean. People may express all kinds of things when circling a higher number on a "person cause" rating scale—perhaps that there was an important cause inside their own skin; or perhaps that they had a particular reason for acting; or perhaps just that the event was intentional. Given the centrality of intentionality in social cognition, it seems likely that one of people's first reactions to a self-relevant event is an intentionality judgment—and for negative events, in particular, the assurance that it was *not* intentional. When actors are asked to explain a relapse in drinking, an aggressive outburst, or a failure on a test, they are apt to communicate right away that these events were unintentional, not a reflection of their own thoughts and plans.[5] And that seems reasonable; a person is unlikely to plan a relapse, decide to have an outburst, or try to bring about failure. Observers, meanwhile, may express something different with their (relatively higher) "person" ratings: they may be less concerned with indicating intentionality but with the fact that surely there was something in the actor that caused the relapse, the outburst, or the failure. Thus, without differentiating the various interpretations of the measures and delineating the distinct interpretations actors and observers adopt, we cannot easily compare actors' and observers' ratings and conclude what their attributions really tell us.

We also need to ask exactly what the evidence is for "biased" actor explanations—usually understood to be explanations that unjustifiably

[5] It appears that social psychologists, too, sometimes treat "situational attributions" as unintentionality judgments. Aronson et al. (2010) cite the following "situational explanation" for someone occupying a seat designated to commemorate Rosa Parks: the person "hadn't seen the sign" (p. 109). The authors consider this a charitable explanation; but its force lies not in anything "situational" (the seeing is actually in the person); the charitable force lies in the fact that the person *unintentionally* hadn't seen the sign.

portray the actor in a positive light. Many studies in which people explained negative events confronted participants with an unusual event, such as an extramarital affair, giving in to an opportunity to cheat or—in almost half of the cases—a failure artificially created by false experimenter feedback (e.g., Chen, Yates, & McGinnies, 1988; Sherrod & Goodman, 1978). Unusual events are likely to contradict an actor's knowledge base. If Audrey generally does well on creativity tests but "learns" that she did worse than most of her peers, it is not self-serving but rather justified to go with the base rates and assume that this particular outcome was a fluke, caused by local, temporary factors (cf. Swann & Read, 1981). If the false feedback in an experiment is sufficiently negative so that most people normally do better than the feedback indicates, the average shift of actors' attribution ratings away from the person category need not imply self-servingness. Observers, by contrast, have no base-rate knowledge that would contradict the (false) negative information they receive about the agent. They have an *N* of 1 that indicates the actor did badly, and when they are pressed for an explanation, observers may be justified in attributing the event at least partially to person factors.

Many of the behaviors studied in the literature on self-servingness are unintentional (and especially the negative ones). The third question is what people actually do when they explain negative *intentional* behavior. Here, the folk-conceptual theory offers a more fine-grained analysis of potential self-protective strategies than the person versus situation model, because the various explanatory tools carry distinct functional capacities. Moreover, we can directly test hypothesized processes that guide people's explanatory choices—information access and impression management. In the case of negative intentional actions, observers most likely have a significant information disadvantage, because such actions are rare, typically violating rather than conforming to a cultural script, and may even occur in unusual circumstances unknown to the observer. Impression management motives will play a significant role as well. Besides the self-protective motive the actor may display, we can make distinct predictions about two kinds of observers—those who try to accuse the agent and those who merely witness and try to make sense of the action. The witness may be willing to take the agent's perspective and search for rational, justifying reasons, and when failing in this attempt, resort to (excusing) CHR explanations. The accuser may jump to (denigrating) CHR explanations right away, denying the agent the subjective, rational viewpoint that could possibly make the action look reasonable.

5.3. Explanations of group behaviors

The third domain of application I would like to discuss is explanations of behaviors performed by group agents. Social psychology has long focused on the social perception of individual agents, but interest has grown in the lay perceptions of groups (Hamilton & Sherman, 1996; Yzerbyt, Judd, & Corneille, 2004). Cognitive scientists, too, have reinvigorated a long-standing debate over whether it makes sense to speak of "group agents" and "group minds." At least it seems to make sense to ordinary people (Bloom & Veres, 1999; Clark, 1994), with some limitations (Knobe & Prinz, 2008). Given that people do apply the same conceptual framework (of intentionality, reasons, etc.) to group agents as they do to individual agents (Malle, 2010; O'Laughlin & Malle, 2002), how do they handle the specific challenges they face with *explaining* group behaviors?

The explainer's challenge is to provide an informative and efficient explanation for a behavior that in a sense is performed by many agents. "Why did so many New Yorkers go to the Kandinsky retrospective at the Guggenheim?" Surely, different New Yorkers had different reasons, and an explainer cannot list all those different reasons, nor should he limit the list to one reason that may apply to only a few of the many. This challenge can be solved with a feature of causal history explanations that I have not highlighted so far: their capacity to integrate multiple distinct reason explanations into one single background explanation. Because CHR explanations precede and bring about reasons, they can sometimes bring about multiple different reasons. Whatever the specific reasons were that various New Yorkers had for going to the exhibit, the explainer might point to a broader background: that the Museum had a wide-ranging campaign months ahead of time, showing colorful Kandinsky paintings in the subway tunnels. That CHR factor may have triggered different reasons in different people, but it also united them to generate the collective action, and so it explains the action more parsimoniously than a long list of different reasons could.

Thus, we predicted (O'Laughlin & Malle, 2002) that people would use a larger number of CHR explanations for behaviors performed by group agents than for the same behaviors performed by individuals. Indeed this was the case. Whereas actions performed by individuals elicited 29% CHR explanations, the same actions performed by groups elicited 44% CHR explanations (Study 1).

However, we proposed that there are two types of group agents. "Many New Yorkers" or "High school seniors nationwide" are *aggregate groups*, in which the members of a group all perform the same action but do so independently. By contrast, in *jointly acting groups*, the group members act *together* as a single agent (e.g., "The Tribeca Art Club went to the Kandinsky retrospective at the Guggenheim"). In the latter case, there is little pressure to use CHR explanations as an integration of multiple individual reasons, because jointly acting groups normally have to converge on their reasons and an intention to act, or else they would not jointly act. As a result, explainers can offer reason explanations for jointly acting groups just like they do for individual agents. For the purposes of explanation, jointly acting groups are like individual agents.

The results supported the predictions (O'Laughlin & Malle, 2002, Study 2): Whereas aggregate groups elicited 38% CHR explanations, jointly acting groups elicited only 19%. Interestingly, the rate for individual agents was slightly higher (29%), opening the intriguing possibility that people consider jointly acting groups as even more "agentic" than individual agents. In fact, they elicited 81% reason explanations, which is even more than actors use on average to explain their own actions. We replicated this pattern in a follow-up study, which showed both the increased rate of CHR explanations for aggregate group behaviors (54%), contrasted with the lower rate of CHR explanations for individual behaviors (34%), but also an even lower rate for jointly acting group behaviors (14%).

For each of the studies, we also classified the free-response explanations into the traditional categories of person versus situation causes, but this variable showed neither an individual-group asymmetry nor a differentiation between the two types of groups.

5.4. Future research

5.4.1. Cross-cultural investigations

Extant research on culture and attribution has focused on the person–situation (or disposition–situation) dichotomy (Choi, Nisbett, & Norenzayan, 1999). Because of the inadequacy of this dichotomy to capture how people explain behavior, there is still much to learn about cultural variations in the nature and cognitive antecedents of behavior explanations. The folk-conceptual theory's distinctions between multiple explanation types offer several avenues of investigation. Here are two examples.

Choi and Nisbett (1998) suggested that the traditional actor–observer asymmetry does not hold among East Asian cultures. However, because we now know that this tendency, if conceptualized in terms of person and situation causes, does not exist among Westerners either, a lack of the asymmetry within collectivist cultures is no longer diagnostic of any cross-cultural differences. By contrast, we have found reliable evidence for three distinct actor–observer asymmetries within the folk-conceptual framework, and we can meaningfully ask whether such evidence replicates in other cultures.

We can also test more specific hypotheses about the psychological basis of potential cultural differences. For example, if Norenzayan and Nisbett (2000) are correct in their proposal that individuals in collectivist cultures generally attend more to "background," these observers should offer more CHR explanations of another person's behavior because such explanations most directly provide the background against which the action in question was performed. A competing hypothesis, however, would predict a larger number of reason explanations in collectivist cultures, if people in these cultures are more concerned with managing other people's "face"

(Ting-Toomey et al., 1991) and if they express this concern by an increase of reason explanations of other people's behavior.

5.4.2. Autism

Research on autism has been highly instructive for understanding the components and operation of the folk theory of mind (Baron-Cohen, Tager-Flusberg, & Cohen, 2000). Similarly, instructive would be an investigation of behavior explanations in autism.

The folk-conceptual theory suggests that reason explanations are often based on actual perspective taking. How often, we do not know. We may be able to gauge this rate by examining autistic individuals' explanations of intentional action. Reasons that can be derived from learned knowledge structures, scripts, schemas, and the like do not require perspective taking and should therefore pose no serious problem for autistic individuals. Reason explanations that actually require perspective taking—for example, because the agent acts in a surprising, script-breaking fashion—should pose a problem for autistic individuals. And those problems are likely to be most severe for belief reasons, given their more idiosyncratic and context-specific contents.

Two hypotheses regarding actor–observer asymmetries can be tested as well. First, autistic individuals presumably engage in practical reasoning about their own actions, so they may be able to report on their own reasons from direct recall (Herrman, 1994). However, autistic individuals do not engage very much in impression management (Baron-Cohen, 1992; Begeer et al., 2008; Nadig, Vivanti, & Ozonoff, 2009). To the extent that the actor–observer asymmetry in reasons has its source in impression management motives, we should see autistic individuals produce a lower rate of reasons in the actor perspective, eliminating the reason asymmetry. To the extent that the asymmetry has its source in information access processes, autistic individuals should produce a standard rate of reasons, thereby upholding the asymmetry.

Second, because autistic individuals find it difficult to grasp other people's mental states, especially beliefs, they may show little to no marker asymmetry because—even if they can identify a plausible belief content (e.g., from action scripts or from the context shared with the actor)—they would not use belief verbs to represent another person *as having* a belief.

5.4.3. Relationships and communication

Attribution concepts have had a considerable impact on the study of relationships, especially the study of dissatisfaction and conflict in marriage (Bradbury & Fincham, 1990). However, in that literature, initial hopes to have discovered attribution as a critical cognitive antecedent to relationship functioning were dashed by serious measurement problems. Selecting and classifying naturally occurring explanations proved challenging because of the "enormous difficulties [we] encountered in trying to code attributions from actual dyadic (marital) interaction using standard attribution dimensions" (Thomas N. Bradbury, personal communication). Lacking an alternative theory of explanation, researchers therefore asked participants to rate preselected behaviors on a variety of attribution dimension (Fincham & Bradbury, 1992). However, disconcerting collinearities afflicted the ratings of intentionality, locus, controllability, responsibility, etc. (Fincham & Bradbury, 1993; Karney, Bradbury, Fincham, & Sullivan, 1994; cf. Anderson, 1983). Such collinearity is not surprising considering the difficult theoretical distinctions people are expected to make—distinctions that not even scholars in the field have entirely agreed on (Shaver, 1996). When people provide many simultaneous ratings about the same stimuli, they are likely to collapse judgments that would normally play distinct psychological roles. For intentional actions, for example, it is not clear what distinct information is conveyed by the indication that the action is intentional, controllable, that the person caused it, or that the person is responsible for it. None of these ratings, moreover, actually capture people's explanations for the specific behaviors.

An alternative approach based on the folk-conceptual theory of explanations would examine naturally occurring behavior explanations (for one's own and the partner's behaviors), either by recording actual conversations or by eliciting reports of past actions and their explanations. Separately within unintentional and intentional behaviors, the various explanation modes and features would be assessed and in turn related to other variables of interest, such as frequency and intensity of conflicts, relationship satisfaction, and communication patterns.

Relatedly, the folk-conceptual theory may advance research in the broader field of human communication (Bazarova & Hancock, 2010). In these contexts, conceptual assumptions, psychological processes, and linguistic expressions converge on particular strategic objectives—such as when people are under pressure for justification (Knight & Rees, 2008) or when they pursue deceptive or hurtful interpersonal goals (Bazarova & Hancock, 2010).

6. The Folk-Conceptual Approach: Some Costs, Many Benefits

6.1. Costs and how to reduce them

Traditional attribution theory had two major attractions. First, the complexity of people's explanations was boiled down to a compact dichotomy of person causes and situation causes. Second, the measurement of such causal attributions was achieved by simple rating scales. If we examine instead the actual explanations people give for human behavior, there is more measurement work to do. These explanations have to be recorded, transcribed, segmented, classified into appropriate folk-conceptual categories, and analyzed as a multivariate set of explanatory tools. Such demands do not permit quick paper-and-pencil studies. Because the folk-conceptual theory of explanation takes naturally occurring explanations as its domain of application, tests of its assumptions and predictions as well as further steps of expansion come at a higher cost in data collection. But such is true for every research program that tries to capture actual behavior as it occurs in the real world. And should not a good portion of social psychological research be committed to this goal?

There is no doubt that classifying explanations into the appropriate folk-conceptual categories is a daunting task. Some researchers, however, have developed approximations of the content coding that may suffice for specific purposes.

Wang, Lignos, Vatsal, and Scassellati (2006) examined people's explanations for a social robot's behavior and devised a scale of "intentional explanations" on the basis of the major modes of explanation people have available. The scale ranged from all the way unintentional (cause explanations—score of 1) to full-blown intentional (reason explanations—score of 4), with a middle ground covered by enabling factors (score of 2) and CHR explanations (score of 3). The latter two implied intentionality but did not highlight the mindful process of reasoning toward an intention.

Levi and Haslam (2005) focused on the major explanation modes of causes, reasons, and CHR explanations, and they instructed their coders to classify explanations into one of these three categories (see also Kiesler et al., 2006). A simplification, yes, but one that reflects an actual conceptual distinction people make and one that has shown predictive power across a variety of phenomena.

Testing specific hypotheses about the social functions of belief reasons and desire reasons may be simplified as well. In a data set with a sufficient number of reason explanations, both mental state verbs and syntactic properties (e.g., for desire reasons: *in order to*, *so that*) can be used to identify specific reason types with text search routines. The only challenge is to identify unmarked belief reasons because, unlike desires, they have no characteristic syntactic properties. If successful, such searches would also allow for tests of the specific roles of belief markers.

A final approach is to first collect people's full range of explanations for a set of stimulus behaviors. Those explanations can then be reduced to a manageable group and presented to a new group of participants as multiple-choice options, which they may evaluate for explanatory quality or relevance (McClure, Densley, Liu, & Allen, 2001). Selection and presentation of these explanatory options can be guided by the critical categories that people are sensitive to, such as reasons versus causal history explanations or belief reasons versus desire reasons.

6.2. A broad benefit: Reconnecting with other disciplines

Despite the methodological demands of the folk-conceptual theory, its payoffs are considerable: It provides a comprehensive theoretical foundation for behavior explanations; clarifies which actor–observer asymmetries do or do not exist; identifies new phenomena such as differences in explanations of group and individual behavior; and points to new directions in research on self-serving explanations, cross-cultural differences, and relationships. And there is a broader benefit as well: The folk-conceptual approach connects, or in some cases reconnects, social psychology with several other disciplines.

First is the integration of social-psychological research with developmental research on children's emerging theory of mind. The latter research tradition has long embraced the idea that human social interaction relies on a set of concepts and cognitive capacities that allow people to grasp others' behavior by means of grasping their minds. It is remarkable that for more than two decades, the developmental and social literatures on social perception had existed in virtual isolation. The dam may have been broken, however, with growing mutual awareness and influence (Epley & Waytz, 2010; Higgins & Pittman, 2008; Malle & Hodges, 2005; Mull & Evans, 2010; Uleman et al., 2008).

A second connection that should have been made before is to the philosophical literature on folk psychology, intentional action, and reason explanations (Greenwood, 1991; Sandis, 2009; Searle, 1983). This tradition anticipated, through theoretical analysis, some of the features of intentional action explanation that have now been empirically demonstrated, and social psychologists will find many additional considerations and testable hypotheses in this literature.

Closely related is the extended debate on folk psychology as an actual *theory* of mind or a process of *simulating* other minds (e.g., Carruthers & Smith, 1996; Stone & Davies, 1995). Developmental researchers have been strongly represented in this debate, but the social-psychological literature—for example, on empathy and perspective taking—has been largely ignored. This may be changing, however, with the influx of social neuroscience research that very much considers empathy, simulation, and mind perception to be part of the same overarching topic (Decety & Ickes, 2009; Singer, 2009). These connections may in turn lead to more sophisticated analyses of the previously found empathy–attribution relation (Gould & Sigall, 1977; Regan & Totten, 1975).

In Artificial Intelligence work, one can observe rapidly increasing interest in "social robots"—computer systems that can actually interact with humans. One approach has been to make the robot appreciate intentionality and be a goal and plan recognizer (e.g., Schmidt, Sridharan, & Goodson, 1978). Another takes advantage of humans' tendency to infer mental states from facial expressions and designs robots that display emotions (Breazeal, 2004).

Social psychology provided important early work on blame and responsibility attributions (Alicke, 1992; Shaver, 1985; Weiner, 1995). A recent surge of research has lifted these topics to the forefront of an interdisciplinary dialog between cognitive science, experimental philosophy, and social psychology. Central here are questions regarding the relation between intentionality and blame and, more generally, the interplay between cognition and emotion in moral judgment (Cushman, 2008; Guglielmo & Malle, 2010a,b; Haidt, 2001; Knobe, 2010).

Less recent but no less important are connections to the text-processing literature, which has highlighted people's ease and speed in inferring goals, plans, and other mental states from narratives (Graesser, Robertson, Lovelace, & Swinehart, 1980; Graesser, Singer, & Trabasso, 1994). Unfortunately, no connections were made at the time with parallel work in social psychology (Smith & Miller, 1983). A recent revival and integration can be found, however, in an expansion of the spontaneous inference literature, which has turned from a sole focus on traits to goals and other social inferences (Hassin, Aarts, & Ferguson, 2005; Malle & Holbrook, 2011).

Finally, two rarely noticed connections are those to linguistics and sociology. On one side, we find work on causatives and linguistic structures that represent mental state terms (Iwata, 1995; Jackendoff & Culicover, 2003). Though research on implicit causality had made a similar connection (Rudolph & Försterling, 1997), the singular focus on the person–situation distinction and a sole reliance on a covariation theory of explanation had limited this otherwise fruitful line of work. On the side of sociology, we find the tradition of *ethnomethodology*, which—like Heider—concerned itself with the "methods that people use for accounting for their own action and those of others" (Hutchby & Wooffitt, 2008, p. 27). Many related theoretical ideas await rediscovery and empirical testing (e.g., Atkinson & Heritage, 1984; Burke, 1945; Schütz, 1967).

7. Dogmas to Give Up

Traditional theories of causal attribution were founded on three dogmatic assumptions: that all psychological events are explained the same way, that people explain these events by identifying causes inside or outside the agent, and that this identification relies on the computation of covariations between causes and their effects. We have to give up all three dogmas if we want to properly account for behavior explanations.

7.1. The events dogma

Evidence against the events dogma is overwhelming, as this chapter should demonstrate. From philosophy to social neuroscience, from developmental psychology to social psychology, data show that humans fundamentally distinguish between intentional and unintentional behavior. Given the deep divide between the two classes of behavior and the complex conceptual assumptions that accompany intentional behaviors, any theory of behavior explanation must track the two classes separately. Once this is done, substantial differences emerge in people's explanations at the conceptual, cognitive, and linguistic levels.

7.2. The person–situation dogma

The differences documented here between explanations of intentional and unintentional behaviors also provide decisive evidence against the person–situation dogma. This dichotomy has been criticized by many, and for a long time (Buss, 1978; Kruglanski, 1975; Locke & Pennington, 1982; Malle, 1999; Malle et al., 2000; White, 1991; Zuckerman, 1978). Nonetheless, absent an alternative theory, the dichotomy has been retained in handbooks, textbooks, and the broader literature. With an alternative theory of explanation now in hand, however, we can safely give up this dogma.

There is nothing wrong in principle with classifying cause explanations (or CHR explanations) into person-related or situation-related factors. This classification, however, has little explanatory value or predictive power. It does not predict actor–observer asymmetries, it does not predict individual-group differences, its meaning is ambiguous in self-serving patterns, and the incremental value over and above intentionality judgments is at best unclear.

Worse yet, the person–situation dichotomy is entirely out of place in reason explanations. It has long been known that puzzling things happen when researchers try to classify reason explanations into the dichotomous person–situation categories (Antaki, 1994; Monson & Snyder, 1977; Ross, 1977). Now we can clarify why and how these puzzles came about—because the person–situation dichotomy fundamentally misrepresents the properties of reason explanations. There is one puzzle in particular that has recurred in the literature and that illustrates this point.

Consider two explanations (Ross, 1977, p. 176): "Jack bought the house *because it was secluded*" and "Jill bought the house *because she wanted privacy*." Ross found it puzzling that the first would traditionally be classified as a "situation cause" and the second as a "person cause," even though the two were only subtly different on their linguistic surface. The person–situation dichotomy was meant to be a fundamental distinction of *causes* people assign to behavior, not of linguistic surface features. But it was such surface features that guided past classifications of free-response explanations into the person–situation scheme (for discussion and evidence, see Malle, 1999, Study 4; Malle et al., 2000, Study 4).

Attribution theory could neither account for nor resolve this puzzle because it arose in the context of reason explanations, which have a different conceptual and linguistic structure from that of cause explanations. The explanation "Jack bought the house *because it was secluded*" does not in fact describe the seclusion as anything like a "situation cause" remotely acting upon Jack and making him buy the house. Rather, the explanation refers to Jack's (unmarked) *belief that* the house was secluded, which was his reason for buying it. The puzzle can be resolved by noting that any unmarked belief reason with situation content looks like a "situation cause" (e.g., "Jack bought the house *because it was secluded*"), but just adding a mental state marker flips it into a "person cause" ("...*because he knew that it was secluded*"). In such cases, the person–situation classification is a result of the linguistic pattern of mental state markers. In other cases, the person–situation classification is a result of differences in reason types, because we can flip the "situational" belief reason above into a "personal" desire reason, such as "Jill bought the house *so she can be on her own*." Applying the person–situation dichotomy to reason explanations is thus arbitrary and meaningless, and it makes us overlook the actual properties by which reasons vary.

7.3. The covariation dogma

The final dogma we must let go is the claim that behavior explanations are based on the process of covariation reasoning. The covariation dogma is toppled by three facts. The first is that the information requirements of covariation reasoning are unrealistic. Behaviors in real life are rarely repeated by other people, by the same person toward other targets, and by the same person in other circumstances.

The second issue is that one important class of explanations is almost never formed on the basis of covariation information, and that is reason explanations (Knobe & Malle, 2002). Actors will often directly recall their reasons, and observers will consult knowledge structures relevant to the particular agent and the particular action. These knowledge structures may include "covariation" information about what actions often go with what motives in the particular culture, but that is very different from wading through actual covariation computations.

Third, there is little evidence that covariation reasoning is spontaneously used in real life. Most attribution reviews note that Kelley's predictions about the use of covariation information have been tested empirically and received reasonable support (e.g., Försterling, 1992; McArthur, 1972). However, these studies were flawed. For one thing, they used dependent measures already framed in terms of the internal/external categories (people were never asked to provide explanations in their own words). More important, participants were always presented with relevant covariation information, and unsurprisingly they made use of it. By contrast, when people have opportunities to spontaneously seek out covariation information, they show little or no interest in it (Ahn, Kalish, Medin, & Gelman, 1995; Garland, Hardy, & Stephenson, 1975; Lalljee, Lamb, Furnham, & Jaspars, 1984).

When is covariation reasoning helpful? When events occur repeatedly, across different contexts, and for many people, thus making Kelley's covariation questions potentially useful. Events of this kind include illnesses, moods and emotions, and successes or failures in repeat circumstances. But even in those cases, the output of a covariation reasoning process does not deliver a concrete explanation. Strictly speaking, Kelley's model just points to a broad category of cause: "something" about the person or about the situation. To identify an *actual* cause, people need to rely on direct observation and on knowledge structures for what specific causes covary with what specific effects. But in most cases, it will be more efficient to consult knowledge structures right away (Abelson & Lalljee, 1988; Schank & Abelson, 1977).

What processes do people rely on to construct specific explanations? Actual data are scarce on this issue, but future research will have to distinguish between the different modes of explanation people engage in (Malle, 2004, chapter 5). Covariation patterns are likely to play a role in causal history explanations and enabling factor explanations, but not in reason explanations, which are likely to rely on projection, knowledge, and perspective taking. Finally, explanations that are designed to meet impression management goals will be constructed in part by considering what the audience knows, wants to know, and would approve of—another form of reasoning by knowledge structures (Slugoski et al., 1993; Turnbull & Slugoski, 1988).

7.4. Orthogonal topics

Briefly, I want to mention two prominent topics in the attribution literature to which the folk-conceptual theory of explanation does not stand in opposition.

7.4.1. Evaluative responses to attributions

Weiner (2001) made the distinction between *attribution* processes, which answer why-questions, and *attributional* processes, which succeed those answers and have consequences for people's emotions and evaluations of the actor. For example, if a person failed because of lack of effort, observers have less sympathy toward the person than if she failed because of lack of ability. Similarly, if a person caused an accident because he was drunk, observers want to punish the person more than if he caused the accident because he had an epileptic seizure. These responses to achievement and moral conduct require a theory of blame attribution, which itself requires the proper integration of concepts such as responsibility, intentionality, controllability, and the roles of justification and excuse.[6] The folk-conceptual theory of behavior explanation is orthogonal to a theory of blame attribution, even though the two share important concepts (i.e., intentionality and justification, as socially acceptable reason explanations). These shared concepts of course constitute people's folk psychology, which underlies both explanations and moral evaluations of behavior (Guglielmo et al., 2009; Malle, 2010).

7.4.2. Discounting

A mainstay of traditional attribution theory has been the phenomenon of discounting. In its original formulation, discounting referred to a situation in which "the role of a given cause in producing a given effect is discounted if other plausible causes are present" (Kelley, 1972, p. 8). Within the traditional attribution framework, the most attractive pairing of plausible causes was person versus situation, because initially the two were seen as directly competing. Subsequently, researchers recognized that the person–situation pair was not competing, and discounting was discovered to be infrequent and highly sensitive to a range of other factors (McClure, 1998). Those factors include extremity and valence of the event, type of measure (e.g., probability of event occurrence vs. explanatory quality), and conversational function. Theoretical derivations of these conditions have been hard to come by and were often circular, such as when explanatory discounting was predicted for "multiple sufficient schemata" but the schemata were identified on the basis of people's explanatory patterns (Fiedler, 1982). More generally, predicting discounting effects from the dependencies among causes as being "alternatives" or "jointly necessary" is hardly informative because it preselects cases that fit the definition of discounting. The question is what predicts those dependencies in the first place.

One noncircular approach was initiated by Leddo, Abelson, and Gross (1984) and continued by McClure and Hilton (1997): Their proposal was to use different *types of explanation* as predictors of dependency, focusing on "preconditions" (often enabling factors) and "goals" (one type of reason explanation). The folk-conceptual theory can help delineate the properties of these and other modes of explanation that may predict discounting patterns. For example, among reason explanations, when the cited desire and beliefs are part of the practical reasoning argument (desire for O and belief that A will lead to O), they are both necessary for the constitution of the relevant intention and will therefore not be discounted. Multiple belief reasons, too, are unlikely to be discounted (unless they are straightforwardly contradictory) because any intentional action relies on numerous beliefs. When multiple desire reasons are jointly mentioned, discounting is possible,

[6] Though originally the distinction between external and internal causes had some prominence in Weiner's theory, he later featured controllability as the driving force of attributional processes (Weiner, 1995).

especially when the reasons differ in social desirability, as observers may discount the more socially desirable motive in favor of the more selfish or "ulterior" motive. Discounting then becomes a function of suspicion (Fein, 1996).

7.4.3. Explanations as mediators of cognition and behavior

One of the deep insights of attribution research has been that explanations mediate social judgment and behavior (Quattrone, 1985). Shifts from intrinsic to extrinsic motivation, self-perception of attitudes, dissonance reduction, self-handicapping, and misattribution of physiological and affective stimuli—such phenomena can be understood only by identifying the role of explanations. However, sometimes the explanations are of intentional behavior, sometimes of unintentional behavior. Because we know that these types of explanations differ conceptually, cognitively, and linguistically, revisiting the mediating role of attribution in light of an expanded theoretical framework may prove highly productive.

8. Epilogue: Overcoming Traditionalism in Science and Textbooks

Aside from social psychology, no other discipline has portrayed humans as interpreting action in terms of person–situation causes or covariation reasoning. And that unique portrayal has lasted for almost 50 years. Despite rapidly growing empirical attribution research between 1970 and 1990, little theoretical progress was made, and the 1960s canonical view is still reported in most textbooks today. Reliance on early literature can be shown across the board in textbook attribution chapters. For example, the average year of publication of cited articles in the four top-selling social psychology books is 1985 (Aronson, 2002; Aronson et al., 2007; Kassin et al., 2008; Myers, 2010). But despite this commitment to early attribution literature, textbooks do not mention the articles from the 1970s and 1980s that undermined the person–situation dichotomy; or the published counterevidence to the actor–observer asymmetry; or the evidence that covariation reasoning rarely occurs in ordinary explanations of behavior.

How could we update Social Psychology textbooks from the 1980s to the twenty-first century? This chapter has offered a road map of such an update; here is a last fast ride through the territory.

1. Heider's major insight should be the starting point: that people distinguish between impersonal causality and personal causality (i.e., intentionality). Evidence for this claim comes from developmental, cognitive, and more recently from social psychological research.

2. From the concept of intentionality, a theory of action explanation can be derived. Substantial support comes from goal-based attribution research, theoretical work in philosophy, and recent work on reason explanations.
3. Not really following Heider, Kelley, and other attribution theorists provided a theory of causal judgment that can be applied to (some classes of) unintentional events. But the lesson here must be that the "naïve scientist" model does not capture well what people really do when they explain behavior, especially intentional behavior.
4. By reference to a small set of psychological processes (e.g., information access, impression management), a number of interesting patterns of explanation can be developed, such as actor–observer asymmetries and individual-group asymmetries. The actor–observer asymmetry also offers a small lesson in philosophy of science—how a long-held belief, on closer inspection, can turn out to be incorrect but better understood when a more detailed theory is adopted.
5. From another one of Heider's insights, outcome attribution work evolved, which focuses on the psychological consequences of certain causal judgments in evaluatively relevant domains, such as achievement (success, failure) and social-moral conduct. From here, only a small step leads to responsibility and blame attributions, as well as the surging topic of moral judgment, where the intentionality concept again plays a central role.
6. An emerging line of work asks how people actually arrive at their explanations—by covariation reasoning, knowledge retrieval, evidence-based inference, or simulation? More is known about how people arrive at another type of inference: that of personality traits. This is where the classic line by Jones and Davis begins and leads all the way to dual-process models of trait inference. However, cautious questions must be raised about the actual prevalence and erroneousness of these inferences.
7. Taking a step back, it is important to point out that attributions are both in the head and in communication. Research has focused on the former, but the latter is the form in which most people encounter attributions: as clarifications and justifications in conversation, following well-understood conversational rules, and reflecting once more the two major determinants of explanatory choice: information access and impression management.

If stories such as this one were retold, we would admit to the dynamics of science—to change over time, insights and misunderstandings, debates resolved, questions yet unanswered. If such stories were retold, our students would see how knowledge evolves in twists and turns; and we as teachers would begin to demand that textbooks, too, change as much as our knowledge.

REFERENCES

Abelson, R. P., & Lalljee, M. (1988). Knowledge structures and causal explanation. In D. J. Hilton (Ed.), *Contemporary science and natural explanation: Commonsense conceptions of causality* (pp. 175–203). New York, NY: New York University Press.

Ahn, W. K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*, 299–352.

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*, 368–378.

Ames, D. R., Knowles, E. D., Morris, M. W., Kalish, C. W., Rosati, A. D., & Gopnik, A. (2001). The social folk theorist: Insights from social and cultural psychology on the contents and contexts of folk theorizing. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 307–330). Cambridge, MA: The MIT Press.

Anderson, C. A. (1983). The causal structure of situations: The generation of plausible causal attributions as a function of type of event situation. *Journal of Experimental Social Psychology*, *19*, 185–203.

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

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

Aronson, E., Wilson, T. D., & Akert, R. M. (2010). *Social psychology* (7th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Astington, J. W. (2001). The paradox of intention: Assessing children's metarepresentational understanding. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 85–103). Cambridge, MA: MIT Press.

Atkinson, J. M., & Heritage, J. (1984). *Structures of social action: Studies in conversation analysis.* Cambridge, England: Cambridge University Press.

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: The MIT Press.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*, 708–717.

Baron-Cohen, S. (1992). The girl who liked to shout in church. In R. Campbell (Ed.), *Mental lives: Case studies in cognition* (pp. 11–23). Oxford: Wiley-Blackwell.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (Eds.), (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). New York: Oxford University Press.

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

Bazarova, N. N., & Hancock, J. T. (2010). From dispositional attributions to behavior motives: The folk-conceptual theory and implications for communication. *Communication Yearbook*, *34*, 63–91.

Begeer, S., Banerjee, R., Lunenburg, P., Meerum Terwogt, M., Stegge, H., & Rieffe, C. (2008). Brief report: Self-presentation of children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*, 1187–1191.

Bertenthal, B. I. (1993). Infants' perception of biomechanical motions: Intrinsic image and knowledge-based constraints. In C. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 175–214). Hillsdale, NJ: Erlbaum.

Block, J., & Funder, D. C. (1986). Social roles and social perception: Individual differences in attribution and error. *Journal of Personality and Social Psychology*, *51*, 1200–1207.

Bloom, P., & Veres, C. (1999). The perceived intentionality of groups. *Cognition*, *71*, B1–B9.

Bradbury, T. N., & Fincham, F. D. (1990). Attributions in marriage: Review and critique. *Psychological Bulletin*, *107*, 3–33.

Breazeal, C. L. (2004). *Designing sociable robots.* Cambridge, MA: MIT Press.

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

Burke, K. (1945). *A grammar of motives.* New York: Prentice-Hall.

Buss, A. R. (1978). Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology*, *36*, 1311–1321.

Bybee, J. L., Perkins, R. D., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world.* Chicago: University of Chicago Press.

Call, J., & Tomasello, M. (1998). Distinguishing intentional from accidental actions in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *Journal of Comparative Psychology*, *112*, 192–206.

Carruthers, P., & Smith, P. K. (Eds.), (1996). *Theories of theories of mind.* Cambridge, England: Cambridge University Press.

Chen, H., Yates, B. T., & McGinnies, E. (1988). Effects of involvement on observers' estimates of consensus, distinctiveness, and consistency. *Personality and Social Psychology Bulletin*, *14*, 468–478.

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.

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin*, *125*, 47–63.

Clark, A. (1994). Beliefs and desires incorporated. *The Journal of Philosophy*, *91*, 404–425.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*, 353–380.

D'Andrade, R. G. (1987). A folk model of the mind. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 112–148). New York, NY: Cambridge University Press.

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

Davidson, D. (1982). Rational animals. *Dialectica*, *36*, 317–327.

Davis, K. E. (2009). An overlooked classic? Review of 'How the mind explains behavior: Folk explanations, meaning, and social interaction'. *The Journal of Social Psychology*, *149*, 131–134.

Deaux, K., & Wrightsman, L. S. (1988). *Social psychology* (5th ed.). Pacific Grove, CA: Brooks/Cole.

Decety, J., & Ickes, W. J. (Eds.), (2009). *The social neuroscience of empathy*. Cambridge, MA: MIT Press.

Dittrich, W. H., & Lea, S. E. G. (1994). Visual perception of intentional motion. *Perception*, *23*, 253–268.

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

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89–98.

Epley, N., & Waytz, A. (2010). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology* (5th ed., pp. 498–541). Hoboken, NJ: John Wiley and Sons, Inc.

Erickson, D. J., & Krull, D. S. (1999). Distinguishing judgments about what from judgments about why: Effects of behavior extremity on correspondent inferences and causal attributions. *Basic and Applied Social Psychology*, *21*, 1–11.

Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology*, *70*, 1164–1184.

Fiedler, K. (1982). Causal schemata: Review and criticism of research on a popular construct. *Journal of Personality and Social Psychology*, *42*, 1001–1013.

Fincham, F. D., & Bradbury, T. N. (1992). Assessing attributions in marriage: The Relationship Attribution Measure. *Journal of Personality and Social Psychology*, *62*, 457–468.

Fincham, F. D., & Bradbury, T. N. (1993). Marital satisfaction, depression, and attributions: A longitudinal analysis. *Journal of Personality and Social Psychology*, *64*, 442–452.

Fiske, S. T. (2008). *Social cognition: From brains to culture* (1st ed.). Boston: McGraw-Hill Higher Education.

Försterling, F. (1992). The Kelley model as an analysis of variance analogy: How far can it be taken? *Journal of Experimental Social Psychology*, *28*, 475–490.

Garland, H., Hardy, A., & Stephenson, L. (1975). Information search as affected by attribution type and response category. *Personality and Social Psychology Bulletin*, *1*, 612–615.

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vols. 1–2, Vol. 1, pp. 89–150). New York, NY: McGraw-Hill.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*, 733–740.

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.

Graesser, A. C., Robertson, S. P., Lovelace, E. R., & Swinehart, D. M. (1980). Answers to why-questions expose the organization of story plot and predict recall of actions. *Journal of Verbal Learning and Verbal Behavior*, *19*, 110–119.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Greenwood, J. D. (Ed.), (1991). *The future of folk psychology: Intentionality and cognitive science*. Cambridge: Cambridge University Press.

Guglielmo, S., & Malle, B. F. (2010a). Enough skill to kill: Intentionality judgments and the moral valence of action. *Cognition*, *117*, 139–150.

Guglielmo, S., & Malle, B. F. (2010b). Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin*, *36*, 1635–1647.

Guglielmo, S., Monroe, A. E., & Malle, B. F. (2009). At the heart of morality lies folk psychology. *Inquiry: An Interdisciplinary Journal of Philosophy*, *52*, 449–466.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*, 336–355.

Harman, G. (1986). Willing and intending. In R. E. Grandy & R. Warner (Eds.), *Philosophical grounds of rationality: Intentions, categories, ends* (pp. 363–380). Oxford, England: Clarendon Press.

Harvey, J. H., Ickes, W. J., & Kidd, R. F. (1976). A conversation with Fritz Heider. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 3–21). Hillsdale, NJ: Erlbaum.

Harvey, J. H., & Tucker, J. A. (1979). On problems with the cause-reason distinction in attribution theory. *Journal of Personality and Social Psychology*, *37*, 1441–1446.

Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, *41*, 129–140.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review*, *51*, 358–374.

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

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*, 243–259.

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, NY: Springer.

Higgins, E. T., & Pittman, T. S. (2008). Motives of the human animal: Comprehending, managing, and sharing inner states. *Annual Review of Psychology*, *59*, 361–385.

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

Hilton, D. J., Smith, R. H., & Kim, S. H. (1995). Processes of causal explanation and dispositional attribution. *Journal of Personality and Social Psychology*, *68*, 377–387.

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

Hockenbury, D. H., & Hockenbury, S. E. (2006). *Psychology* (4th ed.). New York: Worth.

Hutchby, I., & Wooffitt, R. (2008). *Conversation analysis*. Cambridge, England: Polity.

Iwata, S. (1995). The distinctive character of psych-verbs as causatives. *Linguistic Analysis*, *25*, 95–120.

Jackendoff, R., & Culicover, P. W. (2003). The semantic basis of control in English. *Language*, *79*, 517–556.

Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in Cognitive Sciences*, *4*, 22–28.

Johnson, J. T., Jemmott, J. B., & Pettigrew, T. F. (1984). Causal attribution and dispositional inference: Evidence of inconsistent judgments. *Journal of Experimental Social Psychology*, *20*, 567–585.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. *Advances in Experimental Social Psychology*, *2*, 219–266.

Jones, E. E., & McGillis, D. (1976). Correspondent inference and the attribution cube: A comparative reappraisal. In J. H. Harvey, W. J. Ickes, & R. F. Kidds (Eds.), *New directions in attribution research* (Vol. 1, pp. 389–420). Hillsdale, NJ: Erlbaum.

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). New York: General Learning Press.

Karney, B. R., Bradbury, T. N., Fincham, F. D., & Sullivan, K. T. (1994). The role of negative affectivity in the association between attributions and marital satisfaction. *Journal of Personality and Social Psychology*, *66*, 413–424.

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

Kassin, S. M., Fein, S., & Markus, H. (2008). *Social psychology* (7th ed.). Boston: Houghton Mifflin.

Kelley, H. H. (1960). The analysis of common sense. *PsycCRITIQUES*, *5*, 1–3.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–240). Lincoln: University of Nebraska Press.

Kelley, H. H. (1972). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 1–26). Hillsdale, NJ/England: Lawrence Erlbaum Associates, Inc.

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence*. New York: Wiley.

Kiesler, S., Lee, S., & Kramer, A. D. I. (2006). Relationship effects in psychological explanations of nonhuman behavior. *Anthrozoös*, *19*, 335–352.

Knight, L. V., & Rees, C. E. (2008). "Enough is enough, I don't want any audience": Exploring medical students' explanations of consent-related behaviours. *Advances in Health Sciences Education*, *13*, 407–426.

Knobe, J. (2010). Person as scientist, person as moralist. *The Behavioral and Brain Sciences*, *33*, 315–329.

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

Knobe, J., & Prinz, J. (2008). Intuitions about consciousness: Experimental studies. *Phenomenology and the Cognitive Sciences*, 7, 67–83.

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

Kruglanski, A. W. (1977). The place of naive contents in a theory of attribution: Reflections on Calder's and Zuckerman's critiques of the endogenous–exogenous partition. *Personality and Social Psychology Bulletin*, *3*, 592–605.

Kruglanski, A. W. (1979). Causal explanation, teleological explanation: On radical particularism in attribution theory. *Journal of Personality and Social Psychology*, *37*, 1447–1457.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

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

Lalljee, M., Lamb, R., Furnham, A. F., & Jaspars, J. (1984). Explanations and information search: Inductive and hypothesis-testing approaches to arriving at an explanation. *The British Journal of Social Psychology*, *23*, 201–212.

Leddo, J., Abelson, R. P., & Gross, P. H. (1984). Conjunctive explanations: When two reasons are better than one. *Journal of Personality and Social Psychology*, *47*, 933–943.

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.

Lewin, K. (1936). *Principles of topological psychology*. New York: McGraw-Hill.

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, K. D. (2002). Are descriptions of the self more complex than descriptions of others? *Personality and Social Psychology Bulletin*, *28*, 1094–1105.

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. (1998). *F.Ex: Coding scheme for people's folk explanations of behavior*. Latest version 4.5.3 (2010), Retrieved January 2011 from, http://research.clps.brown.edu/SocCogSci/CodingSchemes.html.

Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, *3*, 23–48.

Malle, B. F. (2002a). The relation between language and theory of mind in development and evolution. *The evolution of language out of pre-language* (pp. 265–284). Amsterdam, NL: Benjamins.

Malle, B. F. (2002b). Verbs of interpersonal causality and the folk theory of mind and behavior. In M. Shibatani (Ed.), *The grammar of causation and interpersonal manipulation* (pp. 57–83). Amsterdam, NL: Benjamins.

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

Malle, B. F. (2005a). Folk theory of mind: Conceptual foundations of human social cognition. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious. Oxford series in social cognition and social neuroscience* (pp. 225–255). New York, NY: Oxford University Press.

Malle, B. F. (2005b). Three puzzles of mindreading. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 26–43). New York, NY: Guilford Press.

Malle, B. F. (2006a). Intentionality, morality, and their relationship in human judgment. *Journal of Cognition and Culture*, *6*, 87–112.

Malle, B. F. (2006b). Of windmills and strawmen: Folk assumptions of mind and action. In S. Pockett, W. P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior? An investigation of the nature of volition* (pp. 207–231). Cambridge, MA: MIT Press.

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

Malle, B. F. (2010). The social and moral cognition of group agents. *Journal of Law and Policy*, *20*, 95–136.

Malle, B. F., & Hodges, S. D. (Eds.), (2005). *Other minds: How humans bridge the divide between self and others*. New York, NY: Guilford Press.

Malle, B. F., & Holbrook, J. (2011). *Is there a hierarchy of social inference? Evidence from a new experimental paradigm*. Manuscript in preparation.

Malle, B. F., & Ickes, W. J. (2000). Fritz Heider: Philosopher and psychologist. In G. A. Kimble & M. Wertheimer (Eds.), *Portraits of pioneers in psychology* (Vol. 4, pp. 193–214). Mahwah, NJ: American Psychological Association.

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. (2001). The distinction between desire and intention: A folk-conceptual analysis. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 45–67). Cambridge, MA: The MIT Press.

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

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. (2001). *Intentions and intentionality: Foundations of social cognition*. Cambridge, MA: MIT Press.

Malle, B. F., & Nelson, S. E. (2003). Judging mens rea: The tension between folk concepts and legal concepts of intentionality. *Behavioral Sciences & the Law*, *21*, 563–580.

Malle, B. F., & Pearce, G. E. (2001). Attention to behavioral events during interaction: Two actor-observer gaps and three attempts to close them. *Journal of Personality and Social Psychology*, *81*, 278–294.

McArthur, L. A. (1972). The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology*, *22*, 171–193.

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

McClure, J., Densley, L., Liu, J. H., & Allen, M. (2001). Constraints on equifinality: Goals are good explanations only for controllable outcomes. *The British Journal of Social Psychology*, *40*, 99–115.

McClure, J., & Hilton, D. J. (1997). For you can't always get what you want: When preconditions are better explanations than goals. *The British Journal of Social Psychology*, *36*, 223–240.

Mele, A. R. (1992). *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.

Monson, T. C., & Snyder, M. (1977). 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.

Morris, C. G., & Maisto, A. A. (2006). *Understanding psychology* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Mull, M. S., & Evans, E. M. (2010). Did she mean to do it? Acquiring a folk theory of intentionality. *Journal of Experimental Child Psychology*, *107*, 207–228.

Myers, D. G. (2010). *Social Psychology* (10th ed.). New York: McGraw-Hill.

Nadig, A., Vivanti, G., & Ozonoff, S. (2009). Adaptation of object descriptions to a partner under increasing communicative demands: A comparison of children with and without autism. *Autism Research: Official Journal of the International Society for Autism Research*, *2*, 334–347.

Norenzayan, A., & Nisbett, R. E. (2000). Culture and causal cognition. *Current Directions in Psychological Science*, *9*, 132–135.

Ohtsubo, Y. (2007). Perceived intentionality intensifies blameworthiness of negative behaviors: Blame-praise asymmetry in intensification effect. *Japanese Psychological Research*, *49*, 100–110.

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

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

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, *36*, 1–16.

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology*, *42*, 593–607.

Quattrone, G. A. (1985). On the congruity between internal states and action. *Psychological Bulletin*, *98*, 3–40.

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.

Robins, R. W., Spranca, M. D., & Mendelsohn, G. A. (1996). The actor-observer effect revisited: Effects of individual differences and repeated social interactions on actor and observer attributions. *Journal of Personality and Social Psychology*, *71*, 375–389.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*, 677–688.

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

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York: Academic Press.

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

Rudolph, U., & Försterling, F. (1997). The psychological causality implicit in verbs: A review. *Psychological Bulletin*, *121*, 192–218.

Sande, G. N., Goethals, G. R., & Radloff, C. E. (1988). Perceiving one's own traits and others': The multifaceted self. *Journal of Personality and Social Psychology*, *54*, 13–20.

Sandis, C. (Ed.), (2009). *New essays on the explanation of action*. New York, NY: Palgrave Macmillan.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.

Schmidt, C. F., Sridharan, N. S., & Goodson, J. L. (1978). The plan recognition problem: An intersection of psychology and artificial intelligence. *Artificial Intelligence*, *11*, 45–83.

Schneider, D. J., Hastorf, A. H., & Ellsworth, P. (1979). *Person perception* (2nd ed.). Reading, MA: Addison-Wesley.

Schütz, A. (1967). *The phenomenology of the social world*. Evanston: Northwestern University Press.

Schueler, G. F. (2001). Action explanations: Causes and purposes. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 251–264). Cambridge, MA: MIT Press.

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

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer Verlag.

Shaver, K. G. (1996). Too much of a good thing? *Psychological Inquiry*, *7*, 244–247.

Sheldon, K. M., & Johnson, J. T. (1993). Forms of social awareness: Their frequency and correlates. *Personality and Social Psychology Bulletin*, *19*, 320–330.

Sherrod, D. R., & Goodman, C. L. (1978). Effects of sex-of-observer on female actors' causal attributions for success and failure. *Personality and Social Psychology Bulletin*, *4*, 277–280.

Singer, T. (2009). Understanding others: Brain mechanisms of theory of mind and empathy. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics* (pp. 251–268). London: Academic Press.

Slugoski, B. R., Lalljee, M., Lamb, R., & Ginsburg, G. P. (1993). Attribution in conversational context: Effect of mutual knowledge on explanation-giving. *European Journal of Social Psychology*, *23*, 219–238.

Smith, E. R., & Miller, F. D. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general method. *Journal of Personality and Social Psychology*, *44*, 492–505.

Stone, T., & Davies, M. (Eds.), (1995). *Mental simulation: Evaluations and applications*. Oxford, UK: Blackwell. Readings in mind and language.

Swann, W. B., & Read, S. J. (1981). Acquiring self-knowledge: The search for feedback that fits. *Journal of Personality and Social Psychology*, *41*, 1119–1128.

Ting-Toomey, S., Gao, G., Trubisky, P., Yang, Z., Kim, H. S., Lin, S., et al. (1991). Culture, face maintenance, and styles of handling interpersonal conflict: A study in five cultures. *International Journal of Conflict Management*, *2*, 275–296.

Tomasello, M. (2003). The key is social cognition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 47–57). Cambridge, MA: MIT Press.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*, 239–257.

Turnbull, W., & Slugoski, B. R. (1988). Conversational and linguistic processes in causal attribution. In D. J. Hilton (Ed.), *Contemporary science and natural explanation* (pp. 66–93). Brighton, Sussex: Harvester Press.

Uleman, J. S., Saribay, S. A., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360.

Wang, E., Lignos, C., Vatsal, A., & Scassellati, B. (2006). Effects of head movement on perceptions of humanoid robot behavior. *Proceedings of the ACM conference on human-robot interaction* (pp. 180–193). Salt Lake City, UT.

Watson, D. (1982). The actor and the observer: How are their perceptions of causality divergent? *Psychological Bulletin*, *92*, 682–700.

Weiner, B. (1972). Attribution theory, achievement motivation, and the educational process. *Review of Educational Research*, *42*, 203–215.

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

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

Weiner, B. (2001). Responsibility for social transgressions: An attributional analysis. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 331–344). Cambridge, MA: The MIT Press.

Wellman, H. M., & Bartsch, K. (1994). Before belief: Children's early psychological theory. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 331–354). Hillsdale, NJ/England: Lawrence Erlbaum Associates, Inc.

Wellman, H. M., & Phillips, A. T. (2001). Developing intentional understandings. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 125–148). Cambridge, MA: The MIT Press.

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*, 245–275.

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

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

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

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

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, *18*, 53–57.

Yzerbyt, V., Judd, C. M., & Corneille, O. (2004). *The psychology of group perception: Perceived variability, entitativity, and essentialism.* New York: Psychology Press.

Zelazo, P. D., Astington, J. W., & Olson, D. R. (Eds.), (1999). *Developing theories of intention: Social understanding and self-control.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Zuckerman, M. (1978). Actions and occurrences in Kelley's cube. *Journal of Personality and Social Psychology*, *36*, 647–656.