

3

Attribution Theories: How People Make Sense of Behavior

Bertram F. Malle

In social psychology, the term *attribution* has two primary meanings. The first refers to explanations of behavior (i.e., answers to why questions); the second refers to inferences or ascriptions (e.g., inferring traits from behavior, ascribing blame to a person). What the two meanings have in common is a process of *assigning*: in attribution as explanation, a behavior is assigned to its cause; in attribution as inference, a quality or attribute is assigned to the agent on the basis of an observed behavior. Despite the connection between these phenomena, they have distinct psychological characteristics (Hamilton, 1998; Hilton, Smith, & Kin, 1995; Malle, in press). This chapter will focus on attribution as behavior explanation because it is a far-reaching cognitive and social phenomenon that is embedded in the larger human search for meaning (Malle, 2004).¹

The discussion will begin with the undisputed founder of attribution work, Fritz Heider, then briefly visit Jones and Davis's contribution, and move on to Harold Kelley's theoretical model. Because many excellent reviews of the standard views on these theories are available (see note 1), I will spend relatively little time recounting them. My goal is rather to point out aspects of classic attribution theories that are not generally emphasized, highlight historical misunderstandings, and bring to light theoretical difficulties that have not been adequately addressed. In the second half of the chapter I then introduce an alternative theory of behavior explanations that builds on previous theories but tries to overcome their major difficulties.

Heider's Theory of Attribution

Fritz Heider developed models of attribution for both object perception and person perception. His theory of object perception (first described in Heider, 1920, his dissertation) is rarely cited today, but it serves as the foundation for his later theory of person perception.

Heider attempted to solve one of the core philosophical problems of phenomenology: the relation between sensory information and real objects. That is, he asked how it was possible that humans perceive qualities of *objects in the world* even though all they have are sensations in the mind. Heider argued that real objects shape “media” such as air pressure, light reflections, and sense organs. These media have a considerable degree of variance (for one thing, they reflect many real objects), but the perceptual apparatus reconstructs real objects from their characteristic effects on the media. Heider labeled this reconstruction *attribution* – a process that generates inferences of the relatively invariant qualities of things from the characteristic variance patterns they cause in their media. Perceivers faced with sensory information thus experience perceptual objects as “out there” because they attribute the sensory data to their underlying cause in the world (Heider, 1920).²

After his early work on object perception Heider turned to the domain of social interactions, wondering how people perceive each other in interaction and especially how they make sense of each other’s behavior. Heider proposed that a process of attribution is involved in person perception as well, but he recognized that person perception is more complex than object perception – due to the manifold observational data available and the various causes (e.g., beliefs, desires, emotions, traits) to which these data can be attributed. In addition, it was clear to Heider that persons are very different targets of perception than inanimate objects. Persons are “perceived as action centers and as such can do something to us. They can benefit or harm us intentionally, and we can benefit or harm them. Persons have abilities, wishes and sentiments; they can act purposefully, and can perceive or watch us” (Heider, 1958, p. 21). Note that Heider repeatedly refers to the intentionality of persons, which he considered a core assumption in the conceptual framework that underlies social perception. With the help of such concepts as intentionality and the inference of wishes, purposes, sentiments, and other mental states, Heider argued, perceivers bring order and meaning to the massive stream of behavioral data.

Even though in one sense person perception is like object perception – a process of extracting invariance out of variance – Heider emphasized two distinct features of person perception. The first is that in the social domain, *variance* refers to the agent’s stream of ongoing behavior and *invariance* refers to the inferred perceptions, intentions, motives, traits, and sentiments. Heider (1958) sometimes used the term *disposition* to refer to these invariances, and even though he occasionally referred to traits and abilities when talking about dispositions (e.g., pp. 30, 80), he considered “motives, intentions, sentiments ... the core processes which manifest themselves in overt behavior” (p. 34). It was the agent’s motives that occupied a special role in Heider’s model: “The underlying causes of events, *especially the motives of other persons*, are the invariances of the environment that are relevant to [the perceiver]; they give meaning to what he experiences” (Heider, 1958, p. 81; emphasis added). In social perception, then, Heider’s terms *disposition* and *invariance* referred primarily to mental states.

The second distinct feature of person perception is that when people perform a causal (i.e., attributional) analysis of human behavior, their judgments of causality follow one of two conceptual models (Heider, 1958, chap. 4). The first is a model of *impersonal causality*, applied to unintentional human behaviors (such as sneezing or feeling sad) and physical events (such as waves splashing or leaves falling). The second is a model of *personal causality*,

which is invoked whenever a human agent performs an intentional action (such as cleaning the kitchen or inviting someone to dinner). “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).

Unfortunately, subsequent attribution research misrepresented both of these crucial features of person perception. First, even though mental states make up the majority of what Heider subsumed under dispositional properties (Heider, 1958, pp. 31–34, chap. 4, *passim*), most scholars portrayed Heider’s notion of *disposition* as referring to stable personality factors (i.e., traits, attitudes, or abilities). For example: “Heider began by assuming that just as objects have enduring qualities that determine their appearances, so people have *stable* psychological characteristics that determine their behavior” (Gilbert, 1998, p. 94; emphasis added). This interpretation of dispositions as necessarily stable can be traced to two early sources. The first is Kelley’s (1960) review of Heider’s (1958) book, according to which “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). The second is Jones and Davis’s (1965) influential paper “From acts to dispositions,” in which they used the term *disposition* to refer to character traits and attitudes only. Hence research on *dispositional attribution* became research on trait inferences.

The second misunderstanding was that many scholars mistook Heider’s distinction between personal and impersonal causality – the terms he used to characterize intentional versus unintentional behavior (Heider, 1958, pp. 100–101) – for a distinction between person causes and situation causes. Kelley (1967) famously posited that in all explanations, “the choice is between external attribution and internal ... attribution” (p. 194). Attribution research applied this person–situation dichotomy to all behaviors alike, whether intentional or unintentional, and thereby eliminated Heider’s central concepts of intention, purpose, and motive from later models of social perception.

Ironically, Heider was credited for – and indeed identified with – “discovering” the simple person–situation dichotomy himself:

How do we search for the causal structure of interpersonal events? According to Heider, we do so by reliance upon attributions to the environment (external factors) or to something about the other person (internal factors). (Weary, Edwards, & Riley, 1994, p. 292)

Central to Heider’s entire theoretical position is the proposition that man perceives behavior as being caused, and that the causal locus can be either in the perceiver or in the environment. (Hastorf, Schneider, & Polefka, 1970, p. 63)

But why the perception that Heider proposed a person–situation dichotomy in attribution? One small section of his book appears to have spawned this claim (Heider, 1958, pp. 82–84). There Heider characterizes any “action outcome” (the result of an action, not the action itself) as “dependent upon a combination of effective personal force and effective environmental force” (p. 82). In elaborations of this claim (pp. 83–87), Heider offered a complex picture. He argued that for an action outcome to occur, there needs to be a concomitance of two elements: the agent’s attempt to perform the action (*trying*) and

supporting factors (*can*) that lie in the agent (effort, ability) or in the environment (e.g., opportunity, luck, favorable conditions). *Trying* is the execution of an intention, so Heider stayed true to his analysis of action in terms of intentionality (personal causality). Only for the *can* forces did Heider apply the distinction between person factors and environmental factors, and these *can* forces play a very circumscribed role: they are the necessary elements for an intentional action to be *successful*, the elements that enable the desired outcome to occur (Heider, 1958, p. 109). Such *enabling factor explanations* (Malle, 1999) answer the specific explanatory question of how it was possible that the action outcome was attained (Malle, Knobe, O’Laughlin, Pearce, & Nelson, 2000; McClure & Hilton, 1997), and only for this explanation mode did Heider introduce the distinction between internal and external causal factors. There is no indication in the text that Heider thought people use the internal–external distinction when explaining behavior in general. On the contrary, 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).

The contrast between the two types of explanations – why the action was chosen and what enabled it to succeed – can be illustrated with the following passage from Gilbert (1998, p. 96):

If a pitcher who wishes to retire a batter (motivation) throws a burning fastball (action) directly into the wind (environmental influence), then the observer should conclude that the pitcher has a particularly strong arm (ability). If a batter tries to hit that ball (motivation) but fails (action), then the observer should conclude that the batter lacked coordination (ability) or was blinded by the sun (environmental influence).

The observer’s reasoning here is entirely focused on accounting for successful or failed outcomes; the question of *why* the batter and the pitcher acted as they did is not answered by reference to either arm, wind, or sun. This *why* question is already answered by mentioning the pitcher’s obvious desire to retire the batter and the batter’s obvious desire to hit the ball and get a run. Talking about the arm, wind, and sun, by contrast, answers the question of how the *outcome was attained* (Malle et al., 2000).

In an interview with Bill Ickes (1976, p. 14), Heider explicitly distinguished between these two questions and hence between two types of explanation:

1. the attribution of *outcomes* to causal factors (i.e., enabling factor explanations);
2. the attribution of *intentional actions* to the actor’s motives (i.e., reasons for acting).

Heider himself never developed a model of motive attributions or reason explanations, and he felt that these explanations had not been adequately treated by contemporary attribution work (Ickes, 1976, p. 14). By contrast, Heider felt that outcome attributions (e.g., what made a student fail or succeed on a test) were well developed in Weiner’s work (e.g., Weiner et al., 1972; Weiner, 1986).

The misperception that Heider proposed the external–internal dichotomy as the fundamental dimension of explanation may thus stem from a confounding of outcomes

(explained by enabling factors) and actions (explained by motives or reasons). The following passages from Hastorf et al. (1970) illustrates this confusion:

Presumably the outcomes of action are caused by some combination of personal characteristics and environmental forces [outcome attribution]. The person may have done something because he had to do it ... or because he wanted to do it [action explanation]. (p. 64)

When we infer that the combination of ability and effort was stronger than the external forces, we infer that internal causality was present [outcome attribution]. Only then do we say such things as “he did it because he wanted to” [action explanation]. (p. 89)

In both of these passages, the authors treat two different explanatory questions as if they were one and the same. The judgment whether “he did it because he wanted to” or “because he had to do it” clarifies the agent’s reasons for an action. These reasons can be given even before the agent tries to perform the action because reasons explain the intention, whether or not the intention gets fulfilled. By contrast, the judgments about ability, effort, or external forces clarify how it was possible that the action outcome was attained (an enabling factor explanation). This explanation can only be given after the agent tried to perform the action – if he succeeded – whereas reason explanations track the motivation for the action in the first place.

Thus, whereas Heider has been consistently credited with introducing the person–situation dichotomy in attribution theory, Heider’s actual theory was predicated on the distinction between *personal causality* (which accounts for intentional events) and *impersonal causality* (which accounts for unintentional events) – later recognized as a central element in social cognition (Malle, Moses, & Baldwin, 2001; Zelazo, Astington, & Olson, 1999). Heider’s analysis of personal causality revealed a complex set of mental states centered on intention, and people, when trying to make sense of intentional behavior, attempt to infer those states (such as beliefs, wishes, and sentiments), which were the reasons that guided the actor’s behavior.

Jones and Davis’s Abandoned Theory of Explanation

Jones and Davis (1965) were the first to introduce a theory of dispositional inference, specifying conditions under which a perceiver infers a stable disposition (personality trait or attitude) from an agent’s behavior. But before they introduced this theory, the first few pages of their famous chapter appeared to go in a different direction. They targeted just the issue that Heider had left open: exactly how people explain intentional action by means of motives or reasons. In this short section, entitled “The Naive Explanation of Human Action: Explanation by Attributing Intentions,” the authors attempted to account for “a perceiver’s inferences about what an actor was trying to achieve by a particular action” (p. 222) and the process of finding “sufficient reason why the person acted” (p. 220), because “the perceiver’s explanation comes to a stop when an intention or motive is assigned that has the quality of being reason enough” (p. 220). Despite an apparent plan to present a theory of action explanation by reasons, this was the last that Jones and Davis wrote about action explanations.³ They immediately turned to the conditions under which perceivers infer traits (e.g., arrogance or dominance; see p. 223) from single behavioral events. This theory was refined by Deborah Conliffe in several models of dispositional attribution (e.g., Gilbert, Pelham, & Krull, 1988; Quattrone, 1982), and it inspired research on stereotypes (e.g., Gilbert & Hixon,

1991; Yzerbyt, Rogier, & Fiske, 1998), judgment biases (e.g., Ross, 1977), and impression formation (e.g., Reeder, 1997; Trope, 1986). However, Jones and Davis (1965) wholly left behind the social perceiver's attempt to find the agent's reasons for acting,⁴ and it was up to the next attribution theorists to tackle this question.

Kelley's Theory of Attribution as Causal Judgment

Kelley's (1967) paper on attribution theory in social psychology is generally considered the first systematic and general treatment of lay causal explanations. Kelley's self-ascribed goal in the paper was "to highlight some of the central ideas contained in Heider's theory" (Kelley, 1967, p. 192). Specifically, the two central ideas on which Kelley focused were:

1. In the attribution process "the choice is between external attribution and internal ... attribution" (Kelley, 1967, p. 194).
2. The procedure of arriving at these external or internal attributions is analogous to experimental methodology.⁵

Two questions must be considered here, one historical, one substantive. First, were these two ideas really central to Heider's theory, as Kelley claimed? Second, do the two ideas together provide a strong foundation for a theory of behavior explanation?

As argued earlier, Heider applied the external–internal dichotomy to action outcomes that either succeeded or failed. Explanations of intentional action, by contrast, involve the agent's reasons, and for Heider an external–internal distinction was not useful in this domain. Nonetheless, to support the notion that Heider considered the process of explaining as analogous to experimental methodology (what Kelley termed the covariation principle), Kelley quoted a passage from the very end of Heider's book (Heider, 1958, p. 297), which is itself largely based on the section "Attribution of Desire and Pleasure" (Heider, 1958, pp. 146–160). In this section, however, Heider focused entirely on the attribution of impersonal or unintentional events such as enjoyment. Heider never claimed that all behavioral events are explained by means of covariation. Historically, then, Heider provided at best partial backing for Kelley's two postulates.

Whether Kelley correctly represented Heider or not, the question remains whether Kelley's theory generally accounts for folk explanations of behavior – and not merely for unintentional events. As a starting point, consider the following example that Kelley offers to illustrate the attribution process:

Am I to take my enjoyment of a movie as a basis for an attribution to the movie (that it is intrinsically enjoyable) or for an attribution to myself (that I have a specific kind of desire relevant to movies)? The inference as to where to locate the dispositional properties responsible for the effect is made by interpreting the raw data (the enjoyment) in the context of subsidiary information from experiment-like variations of conditions. (Kelley, 1967, p. 194)

This example features an actor's attempt to explain enjoyment – an unintentional event. Indeed, throughout his chapter Kelley applies this attribution analysis to "effects such as experiences, sensations, or responses" (p. 196) and "impressions" (p. 197), as well as arousal

states and evaluative reactions (pp. 231–232). All of these events are unintentional. Kelley occasionally claimed that his model also extended to the case of “inferring a person’s intentions from knowledge of the consequences of his actions” (p. 196, see also p. 193). However, no theory, empirical data, or examples clarified how this process might work.

Of course, the absence of such clarification is not yet proof that Kelley’s model fails to account for intentional actions. But difficulties emerge quickly when one applies either the person–situation dichotomy or covariation reasoning to intentional actions.

Consider the following scenario:

Having just arrived in the department as a new Assistant Professor, Pauline finds in her mailbox a note that says “Let’s have lunch tomorrow. Faculty club at 12:30? – Fred.” Pauline is a bit surprised. She met Fred W. during her interview, but she wouldn’t have expected him to ask her out for lunch.

Pauline now tries to explain Fred’s (by assumption intentional) action of leaving the note in her mailbox. Kelley’s attribution model would suggest that Pauline’s choice is between a person attribution (something about Fred caused the action) and a situation attribution (something about the circumstances caused the action). But right away, this is a confusing choice. Surely something about Fred must have been present in order for him to put the note in her mailbox: motives, an intention, a fairly controlled movement – all inescapable implications of Fred’s action being intentional. In some basic sense, intentional actions are always caused by the person (D’Andrade, 1987; Heider, 1958; Kruglanski, 1975; Malle & Knobe, 1997). At the same time, the situation had to play some role in Fred’s choice as well – but the situation as *subjectively represented* by Fred. He wouldn’t have put this note in Pauline’s mailbox if he hadn’t *expected* her to check her mailbox and if he hadn’t *thought* about a good time and place for the lunch and had not *hoped* her response to the invitation to be positive. If Pauline knew Fred’s deliberations, she would at once understand and be able to explain why he wrote the note. By contrast, she would not be able to explain it with the obvious conclusion that “something about Fred caused the action” or the vague insight that “something about the situation caused the action.”

Is there a sense in which the “experimental methodology” Kelley has in mind could prove useful? If Pauline engaged in covariation reasoning, she would have to ask the three famous questions about consensus, distinctiveness, and consistency to arrive at a plausible explanation of Fred’s action. If we play it by the books, however, this will generate few answers. Assume that no other faculty member has so far, on Pauline’s first day on the job, left a note in her mailbox (low consensus). What can she conclude from that? Fred may have wanted to welcome her, or go out on a date with her, or discuss some common research ideas with her – there are just too many possibilities. All of these explanations might be labeled “person attributions,” because they are possible goals/desires Fred had when leaving the note. But making a sheer person attribution is uninformative in this case. Pauline does not doubt that Fred had some goal; she rather wonders what goal Fred had.

Similar problems arise with the other covariation questions: has Fred performed this kind of action before? Pauline won’t know, but assuming she finds out that this is the first time Fred did it (low consistency), she learns only that his action has something to do with

her or with this particular point in time, while still not knowing *why* he did it. Finally, consider that Fred has left this kind of note with other people as well (low distinctiveness). In this case, Pauline may conclude that Fred shows some habit, which is also of limited use. She would want to know specifically whether his habit is to invite all new faculty members, or only women, or members of her research area, etc. Systematic collection of such covariation information (if available) may sometimes prove helpful by ruling out some possible explanations of the action in question. For example, if Pauline somehow learns that Fred doesn't act this way with others and that others don't act this way, it can't just be a habit or department culture. But there are still critical steps to go, and in those steps the explainer will not try to choose between "person" and "situation" but rather infer specific goals, beliefs, and the like, that were – in the explainer's assessment – the reasons for the agent's action.

Over the years, Kelley's covariation model or refinements of it have been tested empirically and appeared to receive reasonable support (e.g., Försterling, 1992; McArthur, 1972; Sutton & McClure, 2001). However, these studies suffered from a pair of heavy demands. First, they used person/situation ratings as dependent measures; people were not actually asked to provide natural explanations. Second, the experimenter always presented covariation information to the participants, and unsurprisingly they made use of it. There is no evidence that people spontaneously search for covariation information when trying to explain behavior. Lalljee, Lamb, Furnham, and Jaspars (1984) asked their participants to write down the kind of information they would like to have in order to explain various events, and covariation information was in low demand under these conditions. Similarly, Ahn, Kalish, Medin, and Gelman (1995) allowed people to choose between receiving covariation information or some other information, and explainers were less interested in covariation information than in information about generative forces or mechanisms. Nonetheless, we can safely assume that people occasionally seek out covariation information for such unintentional events as headaches or moods and such outcomes as success or failure. These events occur repeatedly, in different contexts, and for many people, thus making covariation information more readily available. However, as we have seen, covariation reasoning about person and situation causes is ineffectual in the case of explaining intentional actions (Knobe & Malle, 2002).

To summarize, Kelley's (1967) model of attribution contains two core propositions: (a) that attribution is a choice between external and internal causes and (b) that the cognitive procedure by which people arrive at this choice is covariation assessment. Both propositions are problematic. First, the internal–external dimension may be a relevant distinction in explanations of unintentional events, but it does not capture people's explanations of intentional action. Second, covariation assessment is used far less than has been commonly assumed, and it is not at all useful as a method to generate explanations of intentional actions.

Other Attribution Research

The three classic works by Heider, Jones and Davis, and Kelley were not the only important contributions to the study of attribution. In particular, Weiner made seminal contributions to our knowledge of outcome attribution (Weiner, 1986; Weiner et al., 1972). In the domain of achievements, he analyzed the emotions and evaluations people have of others

who succeed or fail. He highlighted the causal dimension of stability as complementing externality–internality and showed that people who failed because of lack of effort (unstable internal) were evaluated more negatively than those who failed because of inability (stable internal). Later Weiner also analyzed other outcomes that happen to people, such as sickness or stigma, and he focused on the dimension of controllability to account for social perceivers' responses to such outcomes (Weiner, 1995). For example, people are more angry at agents who suffer negative outcomes brought about by controllable causes (e.g., illness because of risky behavior) than agents who suffer negative outcomes brought about by uncontrollable causes (e.g., illness because of a genetic precondition).

An overlooked aspect of Weiner's (1986, 1995) work is that it provides considerable evidence against Kelley's theory, because two additional attribution dimensions are needed besides external–internal to account for people's emotional and moral responses to outcomes, and Weiner's predictions were developed without reference to covariation reasoning.

However, like models of dispositional attribution and covariation reasoning, Weiner's models of outcome attribution do not speak to Heider's demands for a theory of action attribution. Folk explanations of intentional action remain to be accounted for.

More in Heider's spirit, 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, so Buss, and attribution theory confounded the two. Buss's (1978) article drew rather negative responses (e.g., Kruglanski, 1979; Harvey & Tucker, 1979), and mainstream attribution theory remained unaffected by his critique. Over the next decade, other scholars launched similar critiques, arguing that reasons are an autonomous form of explanation (Locke & Pennington, 1982) and that attribution theories must incorporate reasons and goals into their conceptual repertoire (Lalljee & Abelson, 1983; Read, 1987; for a review see McClure, 2002). However, such an integration proved difficult, in part because it was not made sufficiently clear why reasons are used to explain intentional behavior in the first place. What makes intentional behaviors so special that they require a unique mode of explanation?

An important contribution to attribution theory emerged in a series of papers on the conversational nature of explanations (Hilton, 1990; Kidd & Amabile, 1981; Turnbull, 1986), which characterized explanations as answers to *why* questions. Such question–answer pairs sometimes occur in people's own heads, but more often they occur as an actual conversational exchange between a questioner and an explainer. This conversational analysis comes with the important implication that, in answering a *why* question, explainers must take into consideration (a) exactly what the questioner finds puzzling (Hilton & Slugoski, 1986; Turnbull, 1986) and (b) what information the questioner already has available (Slugoski, Lalljee, Lamb, & Ginsburg, 1993). In a sense, the explainer anticipates what kinds of possible answers the questioner had in mind when asking the question (Bromberger, 1965).

The insight that explanations are subject to conversational processes was a minor revolution, because it pulled attributions out of their cognitive rabbit hole and highlighted the fundamentally social nature of explanations. However, research into the conversational features of explanations still did not challenge the conceptual apparatus of person–situation causes inherited from Kelley (Norenzayan & Schwarz, 1999).

The preceding overview has pointed to a number of shortcomings in classic and contemporary attribution theories. First, after Heider, the conceptual framework underlying people's explanations was grossly simplified, ignoring such important concepts as intention, goal, and reason. As a result, no adequate accounts of intentional action were offered. Second, covariation analysis was the only psychological process postulated to drive people's construction of explanations, but because it is used much less than was previously assumed, other processes need to be explored. Third, the conversational aspects of explanations, though clearly important, must be integrated with their conceptual and cognitive aspects.

To resolve these problems, we need to first locate behavior explanations in their proper folk-conceptual framework – the network of concepts that people use to make sense of human behavior. Second, we need to specify the psychological processes that enable people to construct explanations. And third, we need to track the communicative and linguistic manifestations of explanations because they are clearly used for a variety of social-interactive goals, and language is the tool that accomplishes these goals. The folk-conceptual theory of behavior explanation (Malle, 1999, 2004) tries to meet all three of these objectives.

The Folk-Conceptual Theory of Behavior Explanation

Folk concepts of mind and behavior

Past attribution theories ascribed to laypeople a simple framework of “effects” (behaviors, outcomes, or events) and “causes,” with the latter falling into personal (dispositional or internal) and situational (external) causes. This framework is incompatible with what we know about children's emerging theory of mind and behavior – the conceptual network that 4–5-year-olds rely on when interpreting human behavior (Perner, 1991; Gopnik & Meltzoff, 1997; Wellman, 1990). In children's theory of mind, we see the importance of a concept of intentionality, of the mind contrasted with observable behaviors, and of specific mental states, such as beliefs and desires, that are used to explain intentional behavior (Bartsch & Wellman, 1995). It is rather unlikely that when they grow up people forget these concepts and distinctions and suddenly begin to explain behavior using a person–situation dichotomy.

Instead, the adult's folk conception of mind and behavior is continuous with the child's framework, with more numerous and more subtle distinctions but essentially the same two-part structure. First, people have perceptual and cognitive systems that filter, group, and integrate certain stimulus inputs into such concepts or categories as *agent*, *intention*, *belief*, and *reason* (D'Andrade, 1987; Kashima, McIntyre, & Clifford, 1998; Leslie, 1995; Malle, 2004; Malle & Knobe, 1997). Second, people make assumptions about these categories and their relationships (Fodor, 1992; Morton, 1996). For example, coordinated movements of agents are classified into the category *intentional action* (Wellman & Phillips, 2001; Woodward, Sommerville, & Guajardo, 2001), and the concept of intentional action relies on the interplay of multiple mental state categories, including *belief* and *desire* (Malle & Knobe, 1997). Intentionality is arguably the core of this framework (Malle et al.,

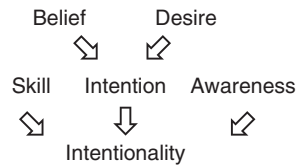


Figure 3.1 A model of the folk concept of intentionality.

Adapted from “The folk concept of intentionality,” by B. F. Malle and J. Knobe, 1997, *Journal of Experimental Social Psychology*, 33, 101–121. © Lawrence Erlbaum Associates, Inc., with permission.

2001). It directly connects behavior with mind by classifying a behavior as intentional when it is characteristically generated by certain mental states (such as belief, intention, and awareness). This centrality of intentionality was already postulated by Heider (1958), but using the different term “personal causality.” This postulate had the least impact on subsequent attribution research, and the greatest weakness of past attribution theories has therefore been the treatment of intentional action. The folk-conceptual theory of explanation tries to address this weakness.

Intentionality. When explaining behavior, people distinguish sharply between intentional and unintentional events (Malle, 1999, 2004, 2007). Social perceivers show a high level of agreement in their intentionality judgments (Malle & Knobe, 1997), and they do so by relying on a shared folk concept of intentionality (Figure 3.1). This concept includes five requirements for an action to be judged as intentional: the action must be based on a desire for an outcome, beliefs about the action’s relationship with this outcome, a resulting intention to perform the action, and skill and awareness when actually performing it (Malle & Knobe, 1997, *Studies* 2–4). Subsequent studies showed that the core concept of intention is seen as a choice and commitment to act that flows from a reasoning process, in which the agent weighs a number of beliefs and desires and settles on a course of action (Malle & Knobe, 2001; Monroe & Malle, 2010).

Modes of explanation. Unintentional events are explained by referring to “mechanical” causal factors (e.g., physical objects and events, but also traits or others’ behaviors), and we may label them *cause explanations* (top of Figure 3.2). That is because people presume no other link mediates between behavior and explanation besides causality (i.e., no components of intentionality such as awareness or intention). Consider the following examples, extracted from original transcripts:

I almost failed my exams. – Why? – Oh, ’cuz I didn’t really prepare for them.
 My dad got mad with me because something was wrong with my computer and he did not know how to fix it.
 A friend cried on the phone. – Why? – She felt that no one loved her.

In contrast, explanations of intentional behavior are far more complex because the folk conception of intentional action is far more complex. As a result, explanations of intentional action break down into three modes (bottom of Figure 3.2): reason explanations, causal history of reason explanations, and enabling factor explanations.

Reason explanations. Reason explanations are the most frequently used mode, comprising about three-quarters of all action explanations (Malle, 2004; Malle, Knobe, & Nelson, 2007). They link directly to the heart of the intentionality concept – the reasoning process leading up to an intention. The concept of intentionality specifies two paradigmatic types of reasons that precede the formation of an intention: the agent’s desire for an outcome and a belief that the intended action leads to that outcome. For example, a student explained why she chose psychology as her major by saying, “I want to go to graduate school in counseling psychology [desire]; I think psychology is the right major to have as background

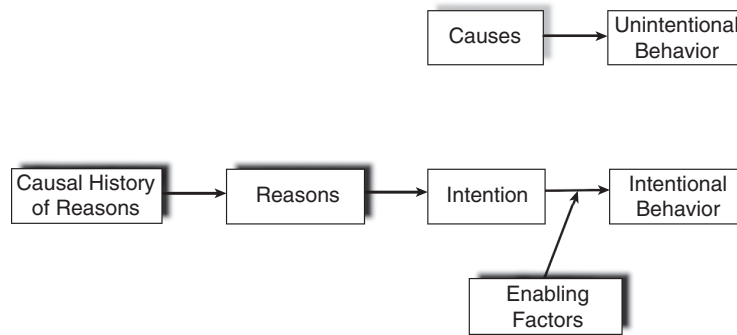


Figure 3.2 Four modes of explanation for unintentional and intentional behavior.

Adapted from “How people explain behavior: A new theoretical framework,” by B. F. Malle, 1999, *Personality and Social Psychology Review*, 3, 23–48. © Lawrence Erlbaum Associates, Inc.

for counseling psychology [belief].” In many naturally occurring explanations, other reasons are mentioned in addition to or instead of the paradigmatic reasons, such as desires for avoiding alternative outcomes, beliefs about the context, beliefs about consequences, and expressions of valuing the action itself.

Reasons have two defining features: subjectivity and rationality. Subjectivity refers to the fact that reason explanations are designed to capture the agent’s subjective reasons for acting. That is, 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. For example, the explanation “She thought she was late for her class” in response to the question “Why did she rush off?” illustrates the subjectivity assumption, because the explainer subtly distances himself from the agent’s belief and implies that, in reality, she probably was not late. But it was that subjective belief, not objective reality, that guided the agent’s action and thus explains it.

Rationality, the second defining feature of reason explanations, refers to the fact that the contents of mental states that are cited as reasons have to hang together so as to offer support for the “reasonableness” of the intention and action they brought about. For an intentional action to be adequately explained by reasons, the action must fulfill the agent’s predominant desire in light of her beliefs. Philosophers often speak of a “practical reasoning argument,” in which reasons are the premises and the decision to act is its conclusion. Consider the following explanation: “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 wanted O [to get in better shape].
- X believed that A [going running] leads to O [getting in better shape].
- Therefore, X decided to A [go running].

In everyday conversation, the logic of rationality will be somewhat cursory and the premises will not always be made explicit. In the earlier example of an agent rushing off, the

action was rationally supported by her belief that she was late for class and 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. Social perceivers would consider it irrational to rush off in this case, or they would assume that some other reason can explain the action in a rational way.

Causal history of reason explanations. Even though people explain most intentional behaviors by reference to the agent's reasons, they explain some of them by pointing to factors that lay in the background of those reasons. These factors can be subsumed under the label *causal history of reasons* (CHR) and include such forces as the agent's unconscious mental states, personality, upbringing, culture, and the immediate context (Malle, 1994, 1999; Malle et al., 2000; O'Laughlin & Malle, 2002). Whereas reason explanations try to capture what the agent herself considered and weighed when deciding to act, causal history explanations take a step back and try to capture what *led up to* the agent's reasons in the first place. For example, when clarifying why Kim didn't vote, an explainer might say "She is lazy" or "Her whole family is apolitical." These statements provide explanations of an intentional action, but they do not pick out Kim's subjective reasons for not voting. Causal history of reason explanations help explain an intentional action by citing causal antecedents to the agent's reasoning and her decision to act, but there is no assumption that the agent actively considered those antecedents in her reasoning process. Hence, when an explainer states that "Kim didn't vote because she is lazy," he does not imply that Kim reasoned: "I am lazy; therefore I shouldn't vote."

Thus, even though CHR explanations help explain intentional actions, they are not subject to the constraints of subjectivity and rationality as reason explanations are (Malle et al., 2000). The agent need not have considered or even been aware of the causal history factors cited in the explanation, nor do CHR factors provide rational support for an explained action. Offering CHR explanations is therefore not an act of perspective taking but rather takes attention away from the element of choice and intention and locates the agent's action in a broader causal nexus.

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 is identified as the critical factor that allowed her to perform the action the way she had intended. Consequently, enabling factor explanations are offered primarily for difficult actions.

The folk-conceptual framework for explaining behavior thus contains four modes of explanation: people explain unintentional behavior with causes, and they explain intentional behavior by either referring to the agent's reasons, citing causal factors that lay in the history of those reasons, or clarifying what factors allowed an agent to successfully perform an intended action. Before we move on to the psychological processes that guide people's

choices among these modes, we need to briefly examine reason explanations more closely, for they have two further features that have shown to have psychological significance: explainers can offer either belief reasons or desire reasons, and they can mark those reasons with appropriate mental state verbs (such as “thought,” “wanted”) or leave them unmarked.

Belief and desire reasons. In people’s folk concept of intentionality, both beliefs and desires serve as necessary conditions of an intention to act (Malle & Knobe, 1997), and both are frequently cited in explanations of intentional action. For example, when explaining why Ian has been working so much lately, one might cite a desire such as “He wants that promotion” or a belief such as “He realizes the project is due in a week.” At times it may not matter whether the explainer mentions a belief reason or a desire reasons because one implies the other (“He thinks hard work will get him the promotion” \Leftrightarrow “He wants that promotion”). But at other times it matters quite a bit. For one thing, belief reasons, more than desire reasons, provide idiosyncratic details about the agent’s decision-making process, such as rejected options, specific plans of action initiation, and considered long-term consequences. For another, belief reasons refer to the agent’s thinking and knowledge, drawing attention to the agent’s rational, deliberative side, whereas desire reasons highlight what the agent wants, needs, hence lacks (Malle et al., 2000).

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

She went to the café because she wants to have an authentic cappuccino.

Or he could use the unmarked form:

She went to the café [____] to have an authentic cappuccino.

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

She went to the café because she thinks they have the best cappuccino.

Or he could use the unmarked belief reason:

She went to the café 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, both for self-presentation and for conveying one’s attitude toward the agent (Malle, 1999; Malle et al., 2000).

In the context of several examples of CHR explanations and reason explanations, Table 3.1 shows both types of reasons, marked and unmarked. (More examples of all modes and types can be found in Malle, 2004). Note that causal history factors can refer to the person,

Table 3.1 Reason explanations and Causal History of Reason (CHR) explanations for three behaviors.

<i>Behavior</i>	<i>Reason explanation</i>	<i>CHR explanation</i>
Kim chose not to vote in the last election.	She thought none of the candidates was trustworthy. [marked belief reason]	She doesn't realize that every vote counts.
	She didn't want to support the system. [marked desire reason]	She is lazy.
Brian used heavy drugs last Sunday at the party.	He was curious what it would feel like. [marked desire reason]	A bunch of others used them.
	He thought it would be cool. [marked belief reason]	He grew up in a drug-dealing home.

Note. The terms *marked* and *unmarked* refer to reason explanations that are expressed either with or without a mental state marker (“he thought,” “she wanted”). Adapted from “How people explain behavior: A new theoretical framework,” by B. F. Malle, 1999, *Personality and Social Psychology Review*, 3, 23–48. © Lawrence Erlbaum Associates, Inc., with permission.

the situation, other people, traits, mental states, or interactions among them. In principle, this choice of “causal locus” could have psychological significance, but in our studies we have found no interesting correlates – neither for CHR explanations nor for cause explanations (e.g., Malle et al., 2007; O’Laughlin & Malle, 2002).

Interlude: measurement concerns

Previously offered alternatives to attribution theory were not able to unseat the traditional theory. This may be in part because those alternatives were not comprehensive enough – focusing, for example, only on goals or only on conversational structures. But another obstacle was researchers’ deeply ingrained habit of using only rating scales to assess attributions. Participants were typically asked to express their explanations on predefined person/situation scales rather than in the more natural form of verbal utterances. As a result, people had to transform their complex explanatory hypotheses into simple ratings, and the fact that they were willing and able to do so was mistaken for evidence that they actually thought in those simple terms.

A few studies analyzed free-response explanations, but the coding was again limited to person–situation categories. This person–situation coding reflects the linguistic surface of explanations, such as the use of mental state markers (McGill, 1989; Nisbett, Caputo, Legant, & Marecek, 1973; for evidence and discussion, see Malle, 1999, Study 4, and Malle et al., 2000, Study 4). Predictable distortions result from treating intentional action explanations within the person-situation frame, illustrated by the following example (cf. Antaki, 1994):

Why are you going Iceland for your holidays? – Because it’s cool there.

The standard attribution treatment of such an explanation would be to call it a “situation cause.” However, the cool weather in Iceland can hardly cause the agent from afar to go

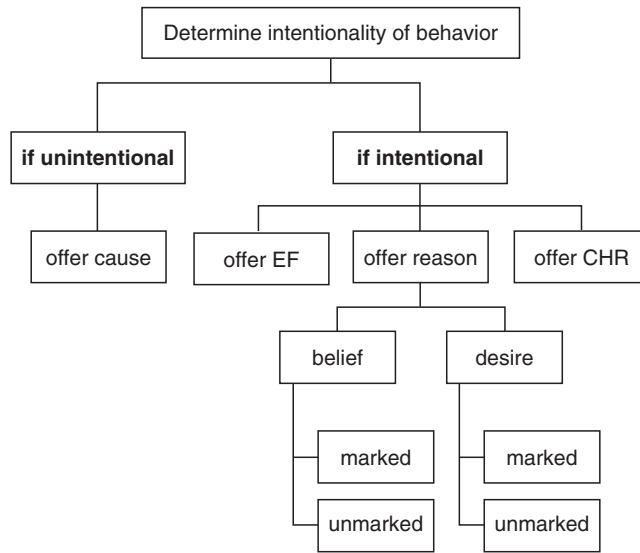


Figure 3.3 Selected choice points in constructing a behavior explanation.

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

there; rather, the agent *thinks* that it is pleasantly cool in Iceland, and that is her (belief) reason for going there. (It would still be her belief reason even if it were in fact warm in Iceland.) Instead of trying to diagnose the locus of some “cause” in the vague space between person and situation, we have to recognize that most folk explanations of intentional behavior refer to the agent’s reasons in light of which and on the grounds of which she acted.

Once the proper folk-conceptual categories (of reasons, causal histories, etc.) are applied to naturally uttered explanations, they can be reliably classified and analyzed (Malle, 1998/2009);⁶ and these classifications show more sensitivity to important psychological variables than either person-situation codings or person-situation ratings (e.g., Malle et al., 2007; O’Laughlin & Malle, 2002). The researcher does the theoretical work here, and participants simply do what they do in everyday life – express explanations of behavior in their own words.

The lesson should be clear: if theoretical progress is to be made in the attribution domain, both theory and measurement must reflect the multifacetedness of people’s explanations. Multiple modes, types, and linguistic forms have to be described, conceptually analyzed, and adequately measured.

Psychological processes underlying behavior explanations

The conceptual layer of the folk-conceptual theory of explanation showed that explainers have a number of choices to make when explaining behavior (Figure 3.3): they need to determine whether the behavior in question is intentional or not; if intentional, whether to explain it with reasons, causal history factors, or enabling factors; and if a reason is chosen, whether to cite the agent’s beliefs or desires and whether to use a mental state marker or not.⁷

Adding a process layer now to the theory, a first important question is what psychological processes determine the explainer's choices among these various explanatory tools. A fair amount of research is available to answer this question, and I will discuss it below. A second process question is how the explainer goes about selecting a specific explanatory content (e.g., the belief reason, "because she thought the door was locked"). No theory of behavior explanation will be able to predict specific contents of explanations, but the psychological processes people rely on when searching for those contents may be identifiable. Currently, very little research is available on this issue, in part because the process of covariation reasoning has been so dominant in the research literature. It is evident, however, that many more processes guide people's search for specific explanation contents – among them knowledge structures, projection, simulation, explicit perspective-taking, and contrast analysis (Ames, 2004; Hilton, 2007; Lalljee & Abelson, 1983; Malle, 2004, 2008). All of these processes deserve greater attention in future research.⁸

What do we know about the psychological determinants of people's explanatory choices? At least three such determinants must be considered:

1. Judgments of behavior attributes
2. Pragmatic goals
3. Information resources

Judgments of behavior attributes. Before offering an explanation, social perceivers make several (often implicit) judgments about the behavior to be explained. One such judgment obviously concerns the *intentionality* of the behavior. If considered unintentional, explanations will refer to a cause; if considered intentional, explanations will cite a reason, causal history, or enabling factor (Malle, 1999).

Among intentional behaviors, a second judgment concerns the *difficulty* of the action. If the action is considered difficult to produce, the explainer will often choose enabling factors; otherwise, he is likely to choose reasons or causal histories (Malle et al., 2000; McClure & Hilton, 1997).

A third judgment determines whether the to-be-explained behavior is singular or represents a trend (across time or across multiple agents). If the behavior is judged to be a trend, the rate of CHR explanations increases (O'Laughlin & Malle, 2002), primarily because multiple instances (or agents) can have multiple reasons, and a CHR explanation can sometimes tie these reasons together by citing an antecedent for all of them (e.g., "Why did you go shopping so many times this week?" – "Because I have three children").

Pragmatic goals. Pragmatic goals refer to the social projects that explainers try to accomplish with their explanations, such as lessen another person's confusion, managing their own status in the interaction, or fending off blame. Two groups of goals can be distinguished by their primary beneficiary: *audience design*, which chiefly benefits the conversation partner, and *impression management*, which chiefly benefits the explainer.

Audience design consists in tailoring an explanation so the audience can learn what it wants to know. The clearest case of such design is when the explainer matches an explanation mode to the type of question asked (Malle et al., 2000; McClure & Hilton, 1998). This question can inquire either about the agent's immediate motivation ("What for?";

which triggers *reasons*), the background of that motivation (“How come?”, which triggers *causal histories of reasons*), or about the factors that enabled successful action performance (“How was this possible?”, which triggers *enabling factors*). More subtle adjustments include offering a belief reason when it can be assumed that the audience already knows the desire reason, or offering a CHR explanation when the reasons for the action are obvious.

Whereas audience design falls under the conversational maxim of relevance (Grice, 1975; Sperber & Wilson, 1986), impression management does not merely try to optimize communication but rather is an act of social influence. Its impact on explanatory choices can be found at several levels of analysis (Malle et al., 2000). People increase their use of reasons, and especially belief reasons, when trying to appear rational; and they explicitly add a mental state marker to their belief reasons when they want to distance themselves from the agent (e.g., “Why is he looking at apartments?” – “He thinks I am moving in with him”).

Information resources. Deficits in relevant information about the agent, the behavior, or the context can limit the explainer’s ability to accomplish his goals, and such deficits often require adjustments of one’s choices among explanatory tools. This can be seen, for example, in the choice of belief reasons vs. desire reasons. We find that observers, compared to actors, offer more desire reasons than belief reasons (Malle et al., 2007). That is because desire reasons are easier to infer from contextual and cultural knowledge, whereas belief reasons are more idiosyncratic and sometimes accessible only to the actor. However, when observers know the agent well or were present when the action took place, their rate of belief reasons increases to resemble that of actors (Malle et al., 2007).

A somewhat more complex case is the role of information access in the choice between reasons and causal history explanations (Malle et al., 2007; O’Laughlin & Malle, 2002). In short, when explaining their own intentional behaviors (as actors), people are in the best position to have every kind of information available and predominantly (in 80% of cases) offer reasons. When explaining other people’s behaviors (as observers), people lack various kinds of information and will try to infer or construct reasons, but this construction reaches a limit, resulting in an average reduction to 65% reasons. Interestingly, this lower rate of reasons persists even when observers are present at the time of action and when observers know the agent well (Malle et al., 2007). This suggests that the actor–observer asymmetry is due not to general knowledge differences but to access differences at a more fundamental level. That is, because actors typically consider their reasons for acting before they decide to act, they have the ability to directly recall their own reasons. Observers, by contrast, must guess those reasons or infer them from observable information.

We can systematically relate the discussed psychological determinants to the range of explanatory choices displayed in Figure 3.2, forming linear equations that connect a particular explanatory choice to a set of psychological determinants (Table 3.2). These equations have received some empirical support (Malle, 1999; Malle et al., 2000; Malle et al., 2007), but more precise relations should emerge from future research.

This theoretical approach of linking the choice of explanatory tools to a limited set of psychological determinants has been successfully applied to predicting strategies of impression management in explanation (Malle et al., 2000), differences between explanations of group and individual behaviors (O’Laughlin & Malle, 2002), and a variety of actor–observer asymmetries in explanations of behavior (Malle et al., 2007). Other domains are open to

Table 3.2 Linking selected explanatory choices to their psychological determinants.

-
- (1) REA/CHR/EF VS. CAU = b_1 (behavior intentionality)
 - (2) EF VS. REA/CHR = b_1 (behavior difficulty) + b_2 (explanatory question) + b_3 (difficulty \times question)
 - (3) CHR VS. REA = b_1 (behavior trend) + b_2 (information) + b_3 (impression management)
 - (4) belief REA vs. desire REA = b_1 (information) + b_2 (impression management)
 - (5) marked belief vs. unmarked belief = b_1 (impression management)
-

Note. REA = Reason explanation. CHR = Causal history of reason explanation. EF = Enabling factor explanation.

investigation as well, such as close relationships, negotiation, psychopathology, and cross-cultural comparisons (see Malle, 2004). To illustrate one application, I describe the death and rebirth of the well-known actor–observer asymmetry in attribution.

Application: actor–observer asymmetries

Few hypotheses have been as widely accepted in the social-psychological literature as Jones and Nisbett's (1972) classic hypothesis of an actor–observer asymmetry in explanations: "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). Surprisingly, a meta-analysis of 173 published studies testing this hypothesis showed that the average effect size for the difference between actor and observer attributions was zero (Malle, 2006). That is, when explanations are analyzed in terms of traditional attribution concepts – referring to either person (or disposition) vs. situation causes – no difference exists between actors' and observers' explanations of behavior.

But do people really explain their own behavior the same way as they explain other people's behavior? They do not. But to see this, we need to adopt a different theory. When explanations are analyzed within the terms of the folk-conceptual theory, strong and reliable actor–observer differences emerge (Malle et al., 2007). In particular, consideration of the choices people make when explaining intentional behavior and the psychological determinants of these choices leads to three actor–observer hypotheses (for more detail, see Malle et al., 2007, pp. 495–496).⁹

First, actors are predicted to use relatively more reason explanations (and fewer CHR explanations) than observers do. That is because actors have better access to their reasons (often through direct recall) and engage in stronger impression management to portray themselves in a positive light (which reasons accomplish better than CHR explanations).

Second, actors are predicted to use relatively more belief reasons (and fewer desire reasons) than observers do. That is because actors' advantage in information access is especially pronounced for the more idiosyncratic and context-specific belief reasons.

Third, actors are predicted to use relatively fewer mental state markers for belief reasons than observers do. That is because actors directly represent the content of their beliefs (e.g., "The plants are dry"), not their own mental state of believing ("I believe: the plants are dry"; Moore, 1993; Rosenthal, 2005). Actors will therefore 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 beliefs* and will therefore typically mark those beliefs (especially when they don’t share them): “Why did she turn the sprinkler on?” – “Because she thought the plants were dry.”

Tests of these predictions across nine studies provided consistent results: Effect sizes for the three novel asymmetries averaged between $d = 0.40$ and $d = 0.69$, whereas effect sizes for the classic asymmetry averaged between $d = -0.08$ and $d = 0.10$. Thus, actors and observers differ in how they explain behavior; but these differences are revealed only once we give up the traditional person–situation framework and instead assess behavior explanations as part of people’s folk theory of mind and behavior – centered on the concept of intentionality and giving rise to multiple modes and types of explanation.

Conclusion

Heider’s (1958) deep insights into the social perception of human action assured attribution phenomena a central position in social psychology. The fundamental assumption that humans spontaneously explain behavioral and social events has led to many insights in the domains of social influence, self-regulation, relationships, and health. However, after 50 years we must acknowledge that traditional formulations of attribution theory either focused too narrowly on inferences of stable traits (following Jones & Davis, 1965) or oversimplified the complex nature of behavior explanations (following Kelley, 1967). Once we identify the conceptual framework within which people actually perceive, interpret, and evaluate behavior, a folk-conceptual theory of behavior explanations emerges that describes and predicts various explanatory phenomena that had previously been overlooked (e.g., reasons, causal history of reasons) or misrepresented (e.g., the actor–observer asymmetry). According to this folk-conceptual theory, people’s explanations of behavior cannot be properly understood when categorized as “person” or “situation” causes. Rather, they fall into multiple distinct modes (causes for unintentional behaviors and reasons, causal histories, and enabling factors for intentional behaviors) and, within modes, into specific types (e.g., belief vs. desire reasons). These are the distinctions that matter when people construct and respond to explanations; and these are the distinctions that must be part of an accurate theory of explanation.

Notes

- 1 For previous reviews of attribution work that gave more weight to trait attributions, see Fiske and Taylor (1991), Gilbert (1998), Hastorf, Schneider, and Polefka (1970), and Ross and Fletcher (1985). For recent integrations of trait attributions within a larger attributional context, see Reeder (2009) and Uleman, Adil Saribay, and Gonzalez (2008).
- 2 Philosophically, this theory may leave much to be desired, but psychologically it was a remarkable model that successfully combined constructivism with causal realism.
- 3 But see Davis (2009), coauthor of the famous Jones and Davis paper, who recently discussed the folk-conceptual theory of explanation.
- 4 In Malle (2004, chap. 1), I discuss Gilbert’s (1998) attempt to find an account of reason explanations in Jones and Davis’s (1965) theory. He pushes such an account as far as one can, but in the best case it captures

goal explanations of limited-option choices and more often it captures only *what* the person was doing, not *why* she was doing it (as Jones and Davis, 1965, themselves point out; p. 222).

- 5 More precisely, “experiment-like variations of conditions [that are] a naïve version of J. S. Mill’s method of difference” (p. 194).
- 6 For applications by other researchers, see Kiesler, Lee, and Kramer (2006); Knight and Reese (2008); and Levi and Haslam (2005).
- 7 There is another layer of choices, which stems from the classic attribution distinctions of person vs. situation and trait vs. nontrait. Causes, enabling factors,

and CHR factors can all be classified into those subtypes. However, our research has not identified any psychological determinants or consequences of these subtypes (e.g., Malle et al., 2007; O’Laughlin & Malle, 2002).

- 8 For illustrations of these processes in naturally occurring explanations, see Malle (2004, pp. 126–145).
- 9 In the explanation of unintentional behavior no asymmetries emerged. Thus, even in the domain in which classic attribution theory is theoretically most applicable (because one can reasonably classify causes into person, situation, and related categories), the predicted asymmetry failed to show.

References

- Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, *54*, 299–352.
- Ames, D. R. (2004). Inside the mind reader’s tool kit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology*, *87*, 340–353.
- Antaki, C. (1994). *Explaining and arguing: The social organization of accounts*. Thousand Oaks, CA: Sage.
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.
- Bromberger, S. (1965). An approach to explanation. In R. Butler (Ed.), *Analytical Philosophy* (2nd series, pp. 72–105). Oxford: Basil Blackwell.
- Buss, A. R. (1978). Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology*, *36*, 1311–1321.
- D’Andrade, R. (1987). A folk model of the mind. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 112–148). New York: Cambridge University Press.
- Davis, K. E. (2009). An overlooked classic? Review of *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction*. (Critical essay). *The Journal of Social Psychology*, *149*, 131–136.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd Ed.). New York: McGraw-Hill.
- Fodor, J. (1992). A theory of the child’s theory of mind. *Cognition*, *44*, 283–296.
- 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.
- Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey, (Eds.), *The handbook of social psychology* (4th edition, pp. 89–150). New York: McGraw Hill.
- Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–210). New York: Guilford.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*, 509–517.
- 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.
- Gopnik, A., & Meltzoff, A. N. (1997). The child’s theory of action. In A. Gopnik & A. N. Meltzoff, *Words, thoughts, and theories* (pp. 125–160). Cambridge, MA: MIT Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.
- Hamilton, D. L. (1998). Dispositional and attributional inferences in person perception. In J. M. Darley & J. Cooper (Eds.), *Attribution and social interaction: The legacy of Edward E. Jones*. Washington, DC: American Psychological Association.
- 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.

- Hastorf, A. H., Schneider, D. J., & Polefka, J. (1970). *Person perception*. Reading, MA: Addison-Wesley.
- Heider, F. (1920). *Zur Subjektivität der Sinnesqualitäten* [On the subjectivity of sense qualities] Unpublished doctoral dissertation. University of Graz, Austria.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, *107*, 65–81.
- Hilton, D. J. (2007). Causal explanation: From social perception to knowledge-based attribution. In A. Kruglanski & E. T. Higgins (Eds.), *Social Psychology: Handbook of basic principles* (2nd Edition). New York: Guilford Press.
- Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, *93*, 75–88.
- Hilton, D. J., Smith, R. H., & Kin, S. H. (1995). Processes of causal explanation and dispositional attribution. *Journal of Personality and Social Psychology*, *68*, 377–387.
- Ickes, W. (1976). A conversation with Fritz Heider. In J. H. Harvey, W. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 3–18). Hillsdale, NJ: Erlbaum.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic Press.
- Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.
- Kashima, Y., 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.
- Kelley, H. H. (1960). The analysis of common sense. A review of “The psychology of interpersonal relations” by Fritz Heider. *Contemporary Psychology*, *5*, 1–3.
- Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 129–238). Lincoln: University of Nebraska Press.
- Kidd, R. F., & Amabile, T. M. (1981). Causal explanations in social interaction: Some dialogues on dialogue. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 3, pp. 307–328). Hillsdale, NJ: Erlbaum.
- Kiesler, S., Lee, S., & Kramer, A. (2006). Relationship effects in psychological explanations of nonhuman behavior. *Anthrozoos*, *19*, 335–352.
- Knight, L.V., & Rees, C. E. (2008). “Enough is enough, I don’t want my audience”: Exploring medical students’ explanations of consent-related behaviours. *Advances in Health Sciences Education*, *13*, 407–426.
- Knobe, J., & Malle, B. F. (2002). Self and other in the explanation of behavior: 30 years later. Special Issue on self-other asymmetries: *Psychologica Belgica*, *42*, 113–130.
- Kruglanski, A. H. (1975). The endogenous-exogenous partition in attribution theory. *Psychological Review*, *82*, 387–406.
- Kruglanski, A. W. (1979). Causal explanation, teleological explanation: On radical particularism in attribution theory. *Journal of Personality and Social Psychology*, *37*, 1447–1457.
- Lalljee, M., & Abelson, R. P. (1983). The organization of explanations. In M. Hewstone (Eds.), *Attribution theory: Social and functional extensions* (pp. 65–80). Oxford: 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. *British Journal of Social Psychology*, *23*, 201–212.
- Leslie, A. M. (1995). A theory of agency. In A. J. Premack, D. Premack, & D. Sperber (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 131–149). Oxford: Clarendon Press.
- 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.
- Locke, D., & Pennington, D. (1982). Reasons and other causes: Their role in attribution processes. *Journal of Personality and Social Psychology*, *42*, 212–223.
- Malle, B. F. (1994). *Intentionality and explanation: A study in the folk theory of behavior*. Unpublished doctoral dissertation, Stanford University, California.
- Malle, B. F. (1998/2009). *F.Ex: Coding scheme for people’s folk explanations of behavior*. University of Oregon. Version 4.5 retrieved June 1, 2009, from <http://www.uoregon.edu/~interact/fex.html> (Original work published 1998)
- Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, *3*, 21–43.
- Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Cambridge, MA: MIT Press.

- Malle, B. F. (2006). The actor-observer asymmetry in causal attribution: A (surprising) meta-analysis. *Psychological Bulletin*, 132, 895–919.
- Malle, B. F. (2007). Attributions as behavior explanations: Toward a new theory. In D. Chadee and J. Hunter (Eds.), *Current themes and perspectives in social psychology* (pp. 3–26). St. Augustine, Trinidad: SOCS, The University of the West Indies.
- Malle, B. F. (2008). The fundamental tools, and possibly universals, of social cognition. In R. Sorrentino and S. Yamaguchi (Eds.), *Handbook of motivation and cognition across cultures* (pp. 267–296). New York: Elsevier/Academic Press.
- Malle, B. F. (in press). Attribution: The search for meaning in human behavior. In N. P. Azari (Ed.), *Encyclopedia of Sciences and Religions*. Springer.
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101–121.
- 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: MIT Press.
- Malle, B. F., Knobe, J., & Nelson, S. (2007). Actor-observer asymmetries in behavior explanations: 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. (Eds.). (2001). *Intentions and intentionality: Foundations of social cognition*. Cambridge, MA: MIT Press.
- McArthur, L. Z. (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. (2002). Goal-based explanations of actions and outcomes. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 12, pp. 201–235). Wiley.
- McClure, J., & Hilton, D. (1997). For you can’t always get what you want: When preconditions are better explanations than goals. *British Journal of Social Psychology*, 36, 223–240.
- McClure, J., & Hilton, D. (1998). Are goals or preconditions better explanations? It depends on the question. *European Journal of Social Psychology*, 28, 897–911.
- McGill, A. L. (1989). Context effects in judgments of causation. *Journal of Personality and Social Psychology*, 57, 189–200.
- Monroe, A. E., & Malle, B. F. (2010). From uncaused will to conscious choice: The need to study, not speculate about, people’s folk concept of free will. *European Review of Philosophy*, 1(3–4).
- Moore, G. E. (1993). Moore’s paradox. In T. Baldwin (Ed.), *G. E. Moore: Selected writings* (pp. 207–212). London: Routledge.
- Morton, A. (1996). Folk psychology is not a predictive device. *Mind*, 105, 119–137.
- Nisbett, R. E., Caputo, C., Legant, P., & Marecek, J. (1973). Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology*, 27, 154–164.
- Norenzayan, A., & Schwarz, N. (1999). Telling what they want to know: Participants tailor causal attributions to researchers’ interests. 1011–1020.
- 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.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology*, 42, 593–607.
- Read, S. J. (1987). Constructing causal scenarios: A knowledge structure approach to causal reasoning. *Journal of Personality and Social Psychology*, 52, 288–302.
- Reeder, G. D. (1997). Dispositional inferences of ability: Content and process. *Journal of Experimental Social Psychology*, 33, 171–189.
- Reeder, G. D. (2009). Mindreading: Judgments about intentionality and motives in dispositional inference. *Psychological Inquiry*, 20, 1–18.
- 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. 174–221). New York: Academic Press.
- Ross, M., & Fletcher, G. J. O. (1985). Attribution and social perception. In G. Lindzey & E. Aronson (Eds.), *The*

- handbook of social psychology* (Vol. 2, pp. 73–114). New York: Random House.
- 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.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.
- Sutton, R. M., & McClure, J. (2001). Covariational influences on goal-based explanation: An integrative model. *Journal of Personality and Social Psychology*, 80, 222–236.
- Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, 93, 239–257.
- Turnbull, W. (1986). Everyday explanation: The pragmatics of puzzle resolution. *Journal for the Theory of Social Behavior*, 16, 141–160.
- Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, 59, 329–360.
- Weary, G., & Edwards, J. A., & Riley, S. (1994). Attribution. In *Encyclopedia of Human Behavior* (Vol. 1, pp. 291–299).
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford.
- Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1972). Perceiving the causes of success and failure. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 95–120). Morristown, NJ: General Learning Press.
- Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.
- 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: MIT Press.
- Woodward, A. L., Sommerville, J. A., & Guajardo, J. J. (2001). How infants make sense of intentional action. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 149–170). Cambridge, MA: MIT Press.
- Yzerbyt, V., Rogier, A., & Fiske, S. T. (1998). Group entitativity and social attribution: On translating situational constraints into stereotypes. *Personality and Social Psychology Bulletin*, 24, 1089–1103.
- Zelazo, P. D., Astington, J. W., & Olson, D. R. (Eds.). (1999). *Developing theories of intention: Social understanding and self-control*. Mahwah, NJ: Erlbaum.