

How People Explain Behavior: A New Theoretical Framework

Bertram F. Malle

*Department of Psychology
University of Oregon*

This article presents a theoretical framework of how people explain behavior. The framework, based on the folk concept of intentionality, distinguishes two major modes of explanation—reason explanation and cause explanation—as well as two minor modes and identifies conditions under which they occur. Three studies provide empirical support for these distinctions. As part of the framework, a detailed model of people’s reason explanations is developed, which emphasizes the unique conceptual and linguistic features of reasons. This model points to limitations of traditional attribution concepts, which are examined theoretically and empirically. Finally, the theoretical framework incorporates attribution concepts, which apply to some but not all modes of explanation. Several paths for future research are outlined—on novel topics such as the roles of rationality and subjectivity in explanations and on classic topics such as the actor-observer asymmetry and the self-serving bias.

Folk explanations of behavior guide people’s perceptions, attitudes, and actions towards each other; they affect impressions, sway sympathies, and alter the paths of relationships. By explaining behavior, people make sense of the social world, adapt to it, and shape it. Behavior explanations are thus themselves a social behavior that must be described and explained.

Attribution theory has been the predominant psychological account of people’s behavior explanations, focusing on the various causes that people assign to behavior (e.g., Heider, 1958; Jones & Davis, 1965; Kelley, 1967, 1973; Weiner, 1986). The picture of ex-

planations that attribution theory has painted, however, differs markedly from the picture that philosophers, developmental psychologists, and sociologists have painted (Audi, 1993; Bartsch & Wellman, 1989; Davidson, 1963; Kalish, 1998; Lennon, 1990; Nichols, 1990; Schult & Wellman, 1997; Scott & Lyman, 1968). These disciplines suggest that people explain intentional behavior differently from unintentional behavior, and these different explanations have typically been called *reason explanations* and *causal explanations*, respectively. Some authors have criticized attribution theory for ignoring the difference between people’s reason explanations and causal explanations (Buss, 1978; Locke & Pennington, 1982; Schneider Hastorf, & Ellsworth, 1982; White, 1991), implying that attribution theory has not delivered an adequate account of how people explain behavior. Despite the gravity of this implication, the cause-reason distinction has not been systematically studied and is surprisingly rarely discussed—never in textbooks and only occasionally in comprehensive reviews (e.g., Antaki, 1994; Fiske & Taylor, 1991).

This article offers an account of the cause-reason distinction within a larger framework of how people explain behavioral events.¹ In short, the framework identifies cause and reason explanations as two among several modes of explanation with which people clarify

Preparation of this article was supported by National Science Foundation CAREER award SBR-9703315.

Portions of this research were described in Bertram F. Malle’s doctoral dissertation completed 1994, Stanford University. Portions were also previously presented in a paper given at the 1994 American Psychological Society symposium “Studying folk psychology: Perceptions of intentionality in primates, children, and adults,” Washington, DC; in a dissertation award address at the 1995 meeting of the Society of Experimental Social Psychology, Washington, DC; in an invited address at the 1997 Western Psychological Association convention, Seattle, WA; and in a paper at the 1997 convention of the Cognitive Science Society, Stanford, CA.

This article has benefited from many critical minds. It has its roots in a term paper for Lee Ross and Mark Lepper’s 1990 graduate social psychology course. Subsequent rewrites were shaped significantly by feedback from Laura Carstensen, Delia Cioffi, Herb Clark, Al Hastorf, Len Horowitz, Shua Knobe, Lara London, Tom Lyon, Ronaldo Mendoza, and H. G. Zilian. Holly Arrow, Ruth Bennett, Garth Fletcher, Sara Hodges, Matt O’Laughlin, and Mick Rothbart provided important comments on more recent drafts of the article.

Requests for reprints should be sent to Bertram F. Malle, Department of Psychology, 1227 University of Oregon, Eugene, OR 97403. E-mail: bfmalle@darkwing.uoregon.edu.

¹ I use the term *behavioral event* (or *behavior*) in a broad way, including observable events (e.g., greeting, crying) as well as unobservable events (e.g., thinking, feeling) and intentional events (e.g., writing, imagining) as well as unintentional events (e.g., stumbling, worrying). Malle and Knobe (1997b) showed that people provide folk explanations for all of these events.

different behaviors and fulfill different goals. In the first part of this paper cause and reason explanations are introduced along with two other modes of explanation, and three studies provide evidence that these modes of explanation are distinct. The second part of the paper develops a detailed model of reason explanations and shows that traditional attribution theory does not adequately describe these explanations. However, attribution concepts are integrated into the larger framework of behavior explanation, which fosters novel research on all modes of explanation and on such classic phenomena as the actor-observer asymmetry and the self-serving bias.

Part 1: A Framework of People's Folk Explanations of Behavior

I begin with two working definitions that are refined shortly: *Reason explanations* are people's explanations of an intentional behavior that cite the agent's reasons for acting that way; *cause explanations* are people's explanations of an unintentional behavior that cite the causes that brought about the behavior. The following sections review past literature on reason and cause explanations and show that the difference between the two explanations derives from the folk concept of intentionality, which leads to precise definitions of reasons and causes.

Reason Explanations and Cause Explanations

Prior discussion of reasons has suffered from two sorts of confusion. First, philosophers have pointed out that reasons are themselves "causes" if we define *causes* broadly as factors that generate an event such as a behavior (Davidson, 1963; cf. Locke & Pennington, 1982). So if reasons are causes and both are generating factors, what good is a distinction between them? The answer is that people regard reasons as a special class of generating factors (those that the agent considers when forming an intention to act) and assign different social, conceptual, and linguistic features to them, as this article demonstrates. To avoid labeling problems, I use the term *generating factors* for the whole class of both reasons and causes and the term *cause* (or *mere cause*) narrowly for those generating factors that are not reasons.

A second source of confusion is that the term *reason* has several meanings in everyday and scientific discourse. The term I target refers to an agent's reasons for choosing to act a certain way. This concept of reason does not include uses that refer to mere causes (e.g., "The reason for our ongoing recession is lack of

consumer confidence"), and it does not include uses that cite evidence and thereby back up claims or beliefs ("There's no reason to believe he's a communist") (cf. Anscombe, 1957; Antaki & Leudar, 1992). People refer to this concept of reason in phrases such as "I think we've got two damned good reasons for getting rid of this now" [LL]² or "What was your reason for choosing Oregon?" Reasons in this sense play a guiding role in the agent's deliberation of whether or not to act, and afterwards they are singled out as the reasons *for which* the agent decided to act. The goal of this section is to clarify exactly what these reasons are, how they can explain behavior, and how they differ from mere causes.

Past Work on the Cause-Reason Distinction

Heider (1958) argued that people use two forms of explanations — one that describes "personal causality" and another that describes "impersonal causality." For Heider, personal causality obtains when "*p* causes *x* intentionally; that is to say, the action is purposive" (p. 100). In this case, "intention ties together the cause-effect relations" (p. 100); the agent "tries to cause *x*, where *x* is his goal" (p. 100), and the answer to "why he is trying to do it" rests on the "reasons behind the intention" (p. 110). Thus, for Heider, reasons (such as wishes; p. 110) explain why an agent forms an intention and tries to act intentionally. Schematically, reasons → intention → intentional action. Heider did not provide any empirical evidence for this model, but we will see later that this is indeed how people conceptualize reason explanations.

Unfortunately, this promising model of reason explanations was overshadowed by Heider's (1958) further analysis. Heider deemed "effects involving persons but not intentions ... as cases of impersonal causality" (p. 101). Schematically, causes → unintentional behavior (without the intervening intention). Inexplicably, however, Heider described the case of impersonal causality as always generated by *situation* causes (e.g., "in the case of impersonal causality, a wide range of environmental conditions will lead to a wide range of effects"; p. 102). This restriction ignores unintentional behavior that is caused by person factors (e.g., my eyes feel heavy because I am tired). Worse

² Most examples of explanations in this paper are transcriptions of actual folk explanations. They were either offered by participants in various studies (usually in response to "why" questions posed by the experimenter) or extracted from spontaneous conversations. Extractions from an electronic version of the London-Lund corpus (Svartvik & Quirk, 1980) are referenced by [LL], and extractions from a sample of teenager conversations (Nolen-Hoeksema, 1994) are referenced by [T].

yet, Heider began to equate impersonal causality with *situation attributions* and personal causality with *person attributions*. Thus, already in Heider's own writing, and more so in subsequent attribution work, the promising distinction between personal causality (reason explanations of intentional behavior) and impersonal causality (cause explanations of unintentional behavior) was collapsed into the simpler distinction between person causes and situation causes. Whereas Heider himself associated person causes with intentional behavior and situation causes with unintentional behavior, subsequent attribution researchers applied the person-situation distinction either only to intentional behaviors (e.g., Jones & Davis, 1965; see Shaver, 1975) or, after Kelley (1967), to all behaviors regardless of their intentionality. Regrettably, the person-situation distinction became Heider's legacy, whereas the distinction between reason explanations of intentional behavior and cause explanations of unintentional behavior never found its way back into mainstream attribution research.

Modern philosophy of action has consistently made this distinction between reason explanations and cause explanations when analyzing people's folk psychology of behavior (Audi, 1993; Lennon, 1990; Mele, 1997). Most authors describe reasons as beliefs or desires that both motivate and rationally support intentional action (e.g., Anscombe, 1957; Davidson, 1963; Goldman, 1970), and they describe intentions as mediating between reasons and action (Brand, 1984; Bratman, 1987; Mele, 1992; Searle, 1983). Thus, philosophy of action arrived at virtually the same schematic as Heider (1958) initially did: reasons (= beliefs, desires) → intentions → intentional actions.

Influenced by this philosophical tradition, Buss (1978, 1979) and Locke and Pennington (1982) attempted to (re-)introduce the distinction between causes and reasons into attribution theory. They characterized reasons, but not causes, as explaining intended action and as making action intelligible and rational. Attribution researchers never disproved these claims, but the lack of empirical evidence in their support may have left researchers skeptical of the cause-reason distinction (e.g., Kruglanski, 1979; McClure, 1984). In addition, the debate around Buss's (1978) initial proposal confounded two very different questions: whether the cause-reason distinction is useful in accounting (a) for the nature of human behavior or (b) for how people explain behavior. Only the second question should concern attribution researchers.

Developmental researchers have been careful to focus on this second question by asking how children learn to explain human behavior. Research has focused on intentional behavior because its explanation requires

an appreciation of the agent's subjective mental states. Studies show that children gradually acquire the concepts of desire and belief between ages 2 and 4 and use them to ascribe mental states to others, such as when explaining their behavior (Bartsch & Wellman, 1989). Some authors call these belief-desire explanations *reason explanations* (Kalish, 1998), others call them *psychological explanations* (Wellman, Hickling, & Schult, 1997). Either way, they refer to the practice of explaining intentional action in terms of the agent's mental states, particularly beliefs and desires. Examples of such explanations given by 3-year-olds include: "Why is Jane looking for the kitten under the piano?" — "She thinks it's under there" or "Why does Andrew put the peanut in his mouth?" — "Because he wants to eat it" (Bartsch & Wellman, 1989).

This brief history of contributions helps refine our working definitions: Cause explanations are people's explanations of unintentional behavior that cite the causes that brought about the behavior without an intervening intention. Reason explanations are people's explanations of intentional behavior that cite the agent's reasons (which can be either beliefs or desires) for acting intentionally, and these reasons lead to intentional action via an intention (see Figure 1).

To sharpen these definitions further, we need to clarify what connection people exactly see between reasons and intentionality. By examining people's folk concept of intentionality we can clarify this connection and highlight the difference between cause and reason explanations.

The Folk Concept of Intentionality

Heider (1958), as noted earlier, offered a conceptual analysis of intentional action; Jones and Davis (1965), Jones and McGillis (1976), and others followed with somewhat different models (for a review see Malle & Knobe, 1997a). None of these models, however, was tested empirically. Malle and Knobe (1997a) provided such an empirical test by examining in detail which conditions need to be fulfilled for people to regard a behavior as intentional. In several studies they supported a model of the folk concept of intentionality that has five conditions: For people to judge an agent's behavior as intentional, the agent must have (a) a desire³ for an outcome, (b) beliefs about a behavior leading to that outcome, (c) a resulting intention to perform that behavior, (d) the skill to perform the behavior, and (e) awareness of fulfilling the intention while performing the behavior.

³ I will adhere to this use of terms: *Desire* refers to the agent's mental state, whereas the content of that mental state (*what* is desired) specifies the agent's *goal* or *outcome*. When the desire is fulfilled, the goal is reached and the outcome obtains.

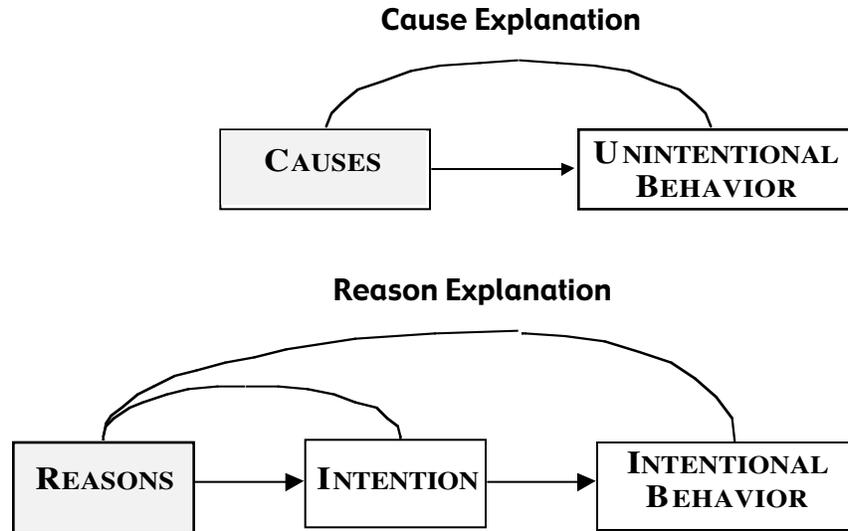


Figure 1. Cause explanations and reason explanations

This folk concept of intentionality specifies how, in people’s conceptualization, an agent brings about an intentional behavior. People’s reason explanations of intentional behavior must therefore refer to at least some of these conditions if they are to clarify what brought the behavior about. I examine each condition in turn, beginning with the ones temporally closest to action and working backwards.

The judged presence of awareness assures that the agent monitors and adjusts her action to fulfill the very intention she has. Awareness, however, has no explanatory function. In response to the question “Why did he leave the room?” the answer “He was aware of it” is not explanatory.

Skill occasionally has explanatory function but not for reason explanations that answer “Why?” questions but rather for explanations that answer “How possible?” questions: “He finished the assignment because he worked hard.” In such cases skill is an “enabling factor”—a factor that enables the intention to be turned into an action and thus clarifies what made it possible for the agent to fulfill the intention (e.g., Heider, 1958, pp. 86-87; Malle, O’Laughlin, & Knobe, 1998; McClure & Hilton, 1997). I say more about enabling factor explanations in a later section.

The intention to act seldom has genuine explanatory value, for the propositional content of an “intention to A” is always the action A itself, leaving it yet to be explained (D’Andrade, 1987): “Why are you going shopping now?”—“Because I intend to go shopping.” A few authors have claimed that intentions are occasionally explanatory, but such claims—often

relying on a very broad concept of intention that includes goals and desires—are debatable.⁴

This leaves the actor’s desire and beliefs, which have been championed as the prototypical reasons in the philosophical and psychological literature since Hume (1740/1978). Not just any belief or desire, however, is by definition a reason; a belief or desire functions as a reason only if the agent considered it in at least a rudimentary reasoning process and formed an intention in light of it. For example, a student explained why she chose psychology as her major: “I want to go to graduate school in counseling psychology. I think psychology is the right major to have as background for counseling psychology.” In light of this desire and this belief, the student chose psychology, and she did it for those reasons.

⁴The claim that intentions are explanatory sometimes stems from a confounding of two meanings of intention—one that has the form *P intends to A* and another that has the form *P acts with an intention to do/have X* (Anscombe, 1957; Bratman, 1987). The latter is best analyzed as “acting with the desire to do/have X” (e.g., Mele, 1992). What appears to be an intention in explanation is thus often a desire reason. For example, “Why did you invite him for lunch?”—“I was trying to be friendly”; in this case the invitation is a means towards the desired end of being friendly. Some authors have argued that intentions can be explanatory when they are themselves reasons for action (Bratman, 1987; Searle, 1991). However, what appears to be an intention in such examples may be analyzed as a belief reason with a past intention as its content. For example, “Why did you invite her for dinner?”—“[I knew] I had promised to meet her to discuss personal issues.”

Citing an intention can be informative in a different manner, namely, as an answer to questions such as, “Now could you tell us about what’s he ah ... trying to do there?” [LL] In answering such questions the intention describes the action at the right level (“I am trying to catch this nasty fly”) without providing its reasons (Anscombe, 1957; Lennon, 1990; Searle, 1983).

For a behavior to be intentional, the agent must have at least one desire (for an outcome) and one belief (that the action leads to that outcome). According to the folk concept of intentionality, this desire and this belief are necessary for an action to be considered intentional. However, this desire and this belief are not necessarily both cited in folk explanations, and other beliefs and desires can be cited as well or instead. This is because an agent who forms an intention does so in light of a broad network of beliefs and desires (Searle, 1983), so the process of reasoning before settling on an intention is presumed to integrate a large number of sometimes incompatible desires, beliefs about consequences, about opportunities to act, and so on. Which of these reasons (and whether one or many) are reported in a given folk explanation depends on the explainer's own familiarity with the agent's cognitive network and on assumptions about the audience's familiarity with it (Hilton, 1990; Kidd & Amabile, 1981; Slugoski, Lalljee, Lamb, & Ginsburg, 1993; Turnbull & Slugoski, 1988).

To summarize, the folk concept of intentionality clarifies the connection between the agent's reasons, intention, and intentionality in people's reason explanations. Reasons are those mental states (primarily beliefs and desires⁵) in light of which the agent formed an intention to act, and these reasons explain the intention as well as the action that fulfills it. Cause explanations refer to factors that are not reasons—factors that bring about a behavior without the mediating role of an intention.

Defining Reasons and Causes

We have now assembled the pieces that allow us to define reasons, causes, and their associated explanations more systematically. A word of caution is in order, however. Defining a folk concept (e.g., reason) aims at reconstructing how people think about and deal with a naturally occurring social phenomenon (e.g., reason explanations). Because people are not perfectly consistent and precise in labeling social phenomena in everyday talk, we may find minor discrepancies between a definition of a folk concept and the particular phenomenon that this definition tries to capture. We cannot circumvent this inherent imprecision by directly defining the "real thing" (as philoso-

⁵A class of mental states that combines features of both beliefs and desires are "valuings" (to like, to enjoy, to fear, etc.). Whether they can be subsumed under the belief or desire category or form a category of their own needs to be determined empirically but is secondary to the present purposes. In the remainder of the article I focus on beliefs and desires and hope that future research will clarify the nature of valuings.

phers sometimes attempt) because the social phenomenon (reason explanations) does not exist independently of being perceived and labeled by people (Searle, 1995). As Heider (1958) has argued so persuasively, to understand people's dealings with a social phenomenon, researchers must understand people's own (folk) concept of this phenomenon (cf. Malle, 1998). If the scientific reconstruction of such a folk concept captures at least the prototype of the phenomenon at issue, we can generate testable explanations and predictions for how people deal with it (in talk, thought, and action), and the scientific endeavor becomes worthwhile.

On to the definitions of concepts that underlie people's folk explanations of behavior:

1. *Generating factors* are states or events that bring about other states or events.
2. *Behavior explanations* cite generating factors of behavioral events (be they intentional or unintentional, observable or unobservable).
3. *Reasons* (as one class of generating factors of behavioral events) are agents' mental states whose content they considered and in light of which⁶ they formed an intention to act.
4. *Mere causes* are those generating factors of behavioral events that are not reasons.
5. *Reason explanations* are those behavior explanations that cite agents' reasons for intending to act or for acting intentionally.
6. *Cause explanations* are those behavior explanations that cite mere causes for an unintentional behavioral event.

If these definitions are adequate reconstructions of people's folk concepts, we should be able to show two basic regularities: First, when people encounter a reason explanation for a behavior, they will be compelled to see the behavior as intentional, but when they encounter a cause explanation for the same behavior, they will be compelled to see it as unintentional (Study 1). This is because, in people's conceptual framework, a reason explanation signals that the explained behavior is intentional, whereas a cause explanation signals that the explained behavior is unintentional. Second, a converse relationship should hold such that when people are asked to explain an unintentional behavior, they will offer cause explana-

⁶I suspect that people do not have a stringent criterion for what it means that the agent *considered* the content of a reason and formed an intention *in light of* it. The prototype of this process is conscious deliberation, in which the agent monitors the strength and importance of desires and actively thinks about facts and their implications; but it may suffice that the agent was dimly aware of certain facts and approved of certain goals when forming an intention. Future research may explore how people deal with the gray area of partly conscious reasons and, perhaps, partly intentional actions (cf. Chan, 1995).

tions, but when they are asked to explain an intentional behavior, they will offer reason explanations (Study 2).

Explanation Mode Predicts Intentionality (Study 1)

A speaker's mode of explaining a given behavior should evoke predictably different judgments of the behavior's intentionality in a listener. According to people's folk theory of behavior, a reason explanation for an agent's behavior implies that the agent intended to perform the behavior, so the behavior should be judged intentional. By contrast, a cause explanation for the behavior implies no intention, so the behavior should be considered unintentional. White (1991) demonstrated that people rate behaviors explained by causes as unintentional and behaviors explained by reasons as intentional. In that study, however, explanations (cause vs. reason) were paired with clearly unintentional vs. intentional behaviors. Participants may have based their intentionality judgments simply on the behavior verbs of the presented descriptions rather than on the explanations themselves. To mend this problem in the present study, I selected behaviors whose intentionality was ambiguous and presented them paired with either a cause explanation or a reason explanation. With behaviors held constant this way, any differences in intentionality ratings must be due to the presence of the cause vs. reason explanations.

Method

Participants. Eighty-six undergraduate students in an introductory psychology class received credit towards a course requirement for participating in the study.

Procedure. A two-page questionnaire was administered as part of a larger survey during a mass testing day. The participants were tested in groups of 5 to 20. The experimenter handed each student a survey booklet with all instructions and measures. After each student had completed the survey, the experimenter thanked and debriefed that student.

Material. From the behaviors studied by Malle and Knobe (1997a), we selected two fairly ambiguous behaviors, which could be construed as either intentional or unintentional: "Anne was driving above the speed limit" and "Vince interrupted his mother." For each behavior, two reason explanations (one citing a

belief, the other citing a desire) and two cause explanations (one citing a person cause and one citing a situation cause) were formulated. For example, the reason explanations for Anne's speeding were: "because she knew that the store closed at 6:00" (belief reason) and "because she wanted to get to the store before 6:00" (desire reason). Its cause explanations were: "because she wasn't paying attention to the speedometer" (person cause) and "because the speedometer didn't work" (situation cause).

Each participant received one of the four explanations of Anne's behavior and one of the four explanations of Vince's behavior. (The order of behaviors was counterbalanced across participants.) After reading each behavior and its explanation, participants rated how intentional that behavior was, using a 9-point scale ranging from 1 (*not at all*) to 9 (*very*).

Results and Discussion

The results support the hypothesis. Anne's behavior was rated as more intentional when explained by a reason ($M = 7.5$) than when explained by a cause ($M = 4.0$), $F(1,83) = 61.4$, $p < .0001$. The results were similar for the belief reason ($M = 7.7$) and the desire reason ($M = 7.2$) as well as for the person cause ($M = 3.9$) and the situation cause ($M = 4.1$). Likewise, Vince's behavior was rated as more intentional when explained by a reason ($M = 8.1$) than when explained by a cause ($M = 5.4$), $F(1,84) = 34.9$, $p < .0001$. The results were again similar for the two reasons ($M = 8.1$ and $M = 8.0$) as well as for the two causes ($M = 5.9$ and $M = 4.9$).

The intentionality ratings for cause explanations were not particularly low, probably because the behaviors were chosen to be ambiguous and thus potentially intentional. Behavior descriptions such as "She was sweating" or "He had a craving for chocolate" would evoke even lower ratings of intentionality, but it would be difficult to pair them with a plausible reason explanation. The next study examined a larger number of behaviors (some clearly intentional, some ambiguous, some clearly unintentional) and tested whether people's explanation mode can be predicted from the intentionality of the explained behavior.

Intentionality Predicts Explanation Mode (Study 2)

This study examined which modes of explanation people use when explaining either unintentional or intentional behaviors. Malle and Knobe (1997a) provided norms of intentionality for 20 behaviors, so we collected explanations for these same behaviors, and the

participants who provided these explanations also rated each behavior for its intentionality. The explanations were then coded for being either reasons or causes.

If the definitions of reason and cause explanations given earlier are correct, behaviors that are rated (on average) as unintentional should be explained by causes and behaviors that are rated (on average) as intentional should be explained by reasons. Moreover, exceptions to this pattern should be clarified by the individual intentionality judgment rendered by the explainer. That is, even if most people offer a cause explanation for, say, grinding one's teeth during the test because they regard it as unintentional, some people may offer a reason explanation ("to appear nervous and tense") if they construe the behavior as intentional. Thus, their individual intentionality rating should be high relative to other people's ratings. Conversely, even if most people offer a reason explanation for, say, interrupting one's mother because they regard it as intentional, some people may offer a cause explanation ("She didn't notice her mother was talking on the phone") if they construe the behavior as unintentional. Thus, their individual intentionality rating should be low relative to other people's ratings.

Method

Participants. Participants were 28 undergraduate students in an introductory psychology class who received credit towards a course requirement.

Material and Procedure. All participants were presented with booklets containing 20 verbally described behaviors. The behaviors were selected to cover a wide range of events, such as bodily states, emotions, actions, and accomplishments.

The participants were first asked to indicate how intentional they thought each behavior was, using an 8-point scale from 0 (*not at all*) to 7 (*completely*). Then they were instructed "to explain, for each of the 20 behaviors, *why* Anne was doing what she did" (observer perspective, $n = 16$) or "to imagine that you did the things described" and "to explain *why* you would do these things" (actor perspective, $n = 12$). Subjects wrote their answers to the "Why?" or "Why would you?" question on two empty lines, the first of which began with the word *Because*.

One behavior, *Anne/You believed that she/you had the flu*, was problematic because people's responses included cause explanations of having the flu ("I got it from my sister"), cause explanations of the belief of having the flu ("I looked at the thermometer and I knew"), and claim backings, that is, evidence from

which the belief could be inferred (e.g., "She was sick with all the symptoms"), which usually are cause explanations as well. Because it was difficult to determine which of these three responses were given, the behavior was excluded from analysis, even though the pattern of results—low intentionality rating and cause explanations—supported the hypothesis.

Coding. Two coders (the author and an undergraduate research assistant) classified the observer explanations, and two coders (the author and a doctoral student) classified the actor explanations according to a comprehensive coding scheme that includes the distinction between reasons and causes (F.EX, 1998; Malle et al., 1998). All coders were blind to the explainers' intentionality judgments. The overall coding agreement was 96% for whether a given response was considered a codeable behavior explanation, and the agreement was 96% ($\kappa = .93$) for the specific mode of explanation (which included reasons, causes, and two minor modes of explanation discussed in the next section). The main analysis below focuses on those 635 explanations (91%) that were reliably classified as either mere causes or reasons; the remaining 9% are discussed afterwards.

Results and Discussion

Malle and Knobe (1997a) reported that people strongly agree in their ratings of intentionality for verbally described behaviors ($\alpha = 0.99$). In the present sample, this agreement among subjects was $\alpha = 0.97$. Moreover, the average ratings in the current sample aligned nicely with the previous intentionality norms ($r = .98$).

Tests of the present hypotheses require measures of the prevalence of each explanation mode for each behavior. One set of prevalence measures comprised (a) the number of explainers who provided at least one cause for a given behavior and (b) the number of explainers who provided at least one reason for that behavior. Because some explainers offered more than one explanation, an alternative set of prevalence measures comprised (a) the total number of cause explanations offered for a given behavior and (b) the total number of reason explanations offered for that behavior. Both types of measures showed that the prevalence of each explanation mode is highly predictable from the explained behavior's intentionality. The intentionality norms correlated at $r = -.92$ with the number of explainers who offered at least one cause explanation and at $r = .92$ with the number of explainers who offered at least one reason explanation (see Table 1). Similar correlations emerged for the to-

Table 1. Relation Between a Behavior’s Perceived Intentionality and the Prevalence of Cause or Reason Explanations

Behavior	Intentionality Norms	Explainers Who offered a	
		Cause	Reason
Won a prize in the lottery	0.94	27	1
Is sweating	1.37	28	
Was yawning during the lecture	1.41	27	2
Was grinding her teeth during the test	2.00	26	
Had a craving for cherries after dinner	2.23	26	
Is in a great mood today	2.70	23	3
Is infatuated with Ben	3.20	25	1
Was worrying about the test	3.69	28	
Got admitted to Princeton	3.78	27	
Interrupted her mother	4.58	3	21
Ignored Greg’s arguments	5.22	3	21
Drove way above the speed limit	5.37	6	19
Applauded the musicians	5.77		25
Greeted her uncle politely	5.94		19
Refused the salesman’s offer	6.22		26
Stole a pound of peaches	6.36		26
Asked Mike out for dinner	6.39		26
Invited Sue to have lunch	6.40		27
Watered her new plants	6.53		26

tal number of cause explanations ($r = -.90$) and the total number of reason explanations ($r = .91$).

There were a few exceptions to the pattern of unintentional behaviors yielding cause explanations and to the pattern of intentional behaviors yielding reason explanations (see Table 1). In seven cases, participants provided a reason explanation for behaviors that were predominantly explained with a cause and on average seen as unintentional. For example, “Imagine you were yawning during the lecture. Why would you?”—“I was trying to hint to the lecturer that s/he was boring.” Or, “Imagine you won a prize in the lottery. Why would you?”—“To get money.” Presumably, subjects who offered these explanations construed the respective behaviors as intentional, and the data tend to confirm this account. In three cases, the explainers who provided these reason explanations had rated the corresponding behaviors much higher on intentionality (5.0, 7.0, 7.0) than the majority who provided cause explanations ($M = 2.5, 1.2, 1.3$, respectively). In two cases, the explainer had rated the corresponding behavior as unintentional and provided a cause but also added a reason, indicating two possible construals (e.g., “Imagine you are in a great mood. Why would you?”—“Something good happened to me or I just decided to be in a good mood”). Only in two cases had the explainers rated the corresponding behavior as unintentional and offered what coders took as a reason (“Imagine you are in a great mood. Why would you?”—“I want to feel good” and “I want to enjoy the day despite anything that might

go wrong”). This small number of inconsistencies may be regarded as measurement error—perhaps the cited desires were not meant as reasons but as non-explanatory comments, so the coders might have been wrong in classifying them as explanations.

The reverse exception refers to cause explanations for behaviors that were typically explained with a reasons and on average seen as intentional; twelve such exceptions occurred (see Table 1). For example, “Anne drove way above the speed limit. Why?”—“Because she didn’t notice the speedometer.” Or, “Imagine you ignored somebody’s argument. Why would you?”—“I was caught up in my own argument.” Presumably, subjects who offered these explanations construed the respective behaviors as unintentional, and the data support this account. The explainers who provided the cause explanations rated the corresponding three behaviors lower on intentionality ($M = 3.0, 3.7, 2.0$) than the majority who provided reason explanations ($M = 5.5, 4.2, 5.2$, respectively). Also note that these exceptions occurred with the three behaviors that were rated lower on the intentionality norms than all other behaviors explained by reasons, indicating subjects’ uncertainty about whether they truly were intentional.

Both types of exceptions indicate that the coding scheme is sensitive to modes of explanations and not just to the coders’ own intentionality judgments. That is, even if coders (along with most other people) regarded a given behavior as intentional, they were still able to code a participant’s cause explana-

tion for it; similarly, even if they saw a given behavior as unintentional, they were still able to code a subject's reason explanation for it. What is critical for a valid coding of folk explanations is recognizing whether the *explainer* saw the behavior as intentional and offered a reason or saw the behavior as unintentional and offered a cause.⁷

As mentioned earlier, 9% of all explanations were coded neither as causes nor as reasons. These were *causal history of reason explanations* (Malle, 1994). For example, "Why would you ignore somebody's arguments? — I am stubborn"; "Why did Anne greet her uncle politely?" — "Because that's how she was taught to." These explanations are given for intentional behaviors, but they cite, not the reasons for which the agent decided to act, but factors that brought about (i.e., lay in the causal history of) these reasons. A more detailed analysis of this explanation mode follows in the next section. It is worth noting, however, that the reliable identification of causal history explanations ($\kappa > .90$) shows once more that the coding scheme is sensitive to modes of explanation and not just intentionality judgments: Causal history of reason explanations emerge when the explainer considers the behavior as intentional but does not cite a reason. A coding scheme that were sensitive merely to the behavior's intentionality would not be able to identify this distinct mode of explanation.

Preliminary Conclusions

Taken together, Studies 1 and 2 provide evidence for a central regularity in people's folk explanations of behavior: Different types of behavior (intentional vs. unintentional) are explained by different modes of explanation. A critic may argue that this regularity is obvious and nothing new. I would agree, but while the regularity has been obvious to philosophers and ordinary folk, most attribution researchers did not differentiate between modes of explanation. The critic may reply that the demonstrated regularity is true by definition and therefore does not constitute an empirical finding (cf. Smedslund, 1988). Again I would agree, but the regularity is true only by folk definition. For example, because people conceptualize ("define") reasons within a framework of intentionality and because this conceptual framework guides their thinking and explaining, people typically explain intentional behavior with reasons. This regularity in

explanation is a genuine social (hence empirical) phenomenon—for if people had no concept of intentionality they would not form reason explanations.

We have seen that people explain intentional behavior with reasons and that the folk concept of intentionality provides the conceptual foundation for these explanations. However, people occasionally provide explanations of intentional behavior that are not reason explanations. The next section introduces these "minor" (less frequent) modes of explanation and shows that the folk concept of intentionality provides the conceptual foundation for these modes as well.

Two Minor Modes Of Folk Explanation

Two further modes of explanation complement reason explanations in clarifying various aspects of intentional behavior. According to the folk concept of intentionality, intentional behavior is caused by the agent's intention, and the agent forms an intention on the basis of reasons. So the major explanatory link to be established for intentional action is that between the agent's reasons and the intention to act, and this link is expressed in reason explanations (see Figure 2). In addition, the structure of the folk concept of intentionality allows for two more explanatory links. One clarifies (via enabling factors) how it was possible that the intention in fact led to the action; the other clarifies (via the causal history of reasons) which factors led to the agent's reasons. I briefly discuss enabling factor explanations (see also Malle et al., 1998; McClure & Hilton, 1997) and then turn to a more extended presentation of causal history of reason explanations.

Enabling Factor Explanations

Intentions are not always carried out. Sometimes an opportunity is absent, the agent lacks skill, or obstacles emerge. When an intentional action did get performed despite such unfavorable conditions, perceivers may require an explanation for how it was possible that the action was in fact performed. For example, "How come John aced the exam?"—"He's a stats whiz." Such *enabling factor explanations* refer to the agent's skill, efforts, opportunities, or to removed obstacles (see McClure & Hilton, 1997; Turnbull, 1986). These explanations take it for granted that the agent had an intention (and reasons) to perform the behavior; what they try to clarify is how it was possible that this behavior was in fact performed.

Because enabling factor explanations clarify how it was possible that the agent performed the action, not

⁷The present results leave open whether people treat intentionality as a dichotomous attribute (suggested by the nearly discrete distribution of reason explanations) or as a continuous one (suggested by the monotonic increase of average ratings). The average ratings are inconclusive, however, for they may represent, not a genuinely continuous judgment, but varying confidence (induced by ambiguous stimuli) in what is normally a dichotomous judgment. Future research should address this issue.

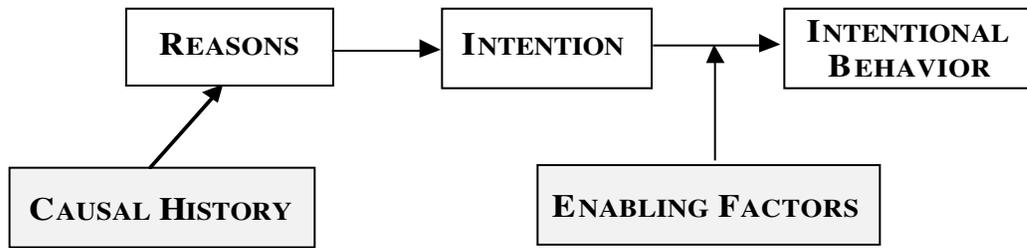


Figure 2. Causal history of reason explanations and enabling factor explanations

what motivated the action, they should be frequent when the explainer directly answers the question “How was this possible?” but should be rare in response to the motivationally biased questions “Why?” or “What for?” Malle et al. (1998) supported this hypothesis, finding that enabling factor explanations occur 4 to 12 times more frequently in response to a “How possible?” question than in response to any other explanatory question. Moreover, enabling factor explanations should be frequent when the behavior was difficult to perform but rare when the behavior was easy (McClure & Hilton, 1997). Malle et al. (1998) supported this hypothesis as well, finding that enabling factor explanations occur 7-8 times more frequently with difficult than with easy behaviors.

The notion of enabling factor explanations helps clarify why explanations of success or achievements never trigger reasons (“How come she got the only A?”—“She is the smartest”), even though they are often intentionally brought about by the agent. Achievements such as acing an exam or finishing a book are presumed to be desirable for the agent, rendering the reasons for trying to achieve them quite obvious. What requires clarification is rather how this achievement was possible, inviting enabling factor explanations.

Causal History of Reason Explanations

Causal history of reason explanations provide an explanatory link between reasons and their own causal history, citing factors that preceded and thus brought about the reasons for an action. These explanations literally describe the *causal history* of reasons, which could lie in childhood, in cultural training, in personality traits, or in a situational cue that triggered a particular desire. Thus, causal history factors offer the context, background, and origin of reasons without denying that the reasons themselves motivated the action. This background function is illustrated in the following diary excerpt in which the writer explains why she is trying to write more neatly: “I have been looking over my writings [causal history] and I can barely understand them [belief reason, caused by the looking]. That’s why I’m trying to write more neatly [intended

action].” It would be sufficient to explain “I’m trying to write more neatly” by saying “I can barely understand my writing”; but the causal history of this reason clarifies how this issue came up in the first place.

One obvious origin of reasons is the agent’s personality, particularly stable traits, as in the following two examples [T]: (a) “Why did Eric vandalize the school gym?”—“He’s just plain reckless.” (b) “Then why would Tanya come up and talk to us out of her own free will?”—“Well ... weird people do these kinds of things.” In both cases, the explainer quite likely perceived the explained behavior as intentional but offered traits to explain the background of the agent’s reasons rather than the reasons themselves (presumably because the explainer did not know those reasons).

Besides traits, behaviors too can function as causal histories of reasons. “I invited my friend to dinner because I wanted to talk to him since *we haven’t talked all quarter*.” The person’s desire (to talk to him) explains (as a reason) why he invited his friend, while the fact that he hadn’t seen him all quarter explains (as a causal history of the reason) why he had the desire to invite him. Similarly, consider a child’s explanation for why lovers hold hands: “Because they want to make sure their rings don’t fall off [reason] because they paid good money for them [causal history].” Sometimes cognitive states can provide the causal history of reasons, as in “Why did she vote against the social security bill?”—“Because she doesn’t understand the needs elderly people have.” Clearly, the politician’s lack of understanding is not her *reason* to vote against the bill but rather a factor that led her to whatever reasons she had. Finally, situational factors can provide a causal history for reasons, as in “She gave me a call because a friend had left my number or something sitting on her desk [causal history], and that’s the only person she could think of to call [reason].” Apparently, the number lying on the desk triggered the belief reason that the agent had for calling the speaker.

Causal history explanation can be found in criminal cases when experiences of childhood abuse are brought forth as mitigating factors in the defense of an apparently intentional homicide, such as in the Me-

nendez case (Wilson, 1997). Causal history explanations can also be found in sociological writings, such as in Durkheim's (1897/1951) famous explanation of suicide, which referred not to the individual's reasons to commit suicide, but to the broader determinants (social isolation) of the various reasons individuals may have. Finally, causal history explanations may assume a psychodynamic flavor, such as when the act of making a joke during an intimate conversation is examined, not for its immediate reasons (e.g., the person wanted to entertain his partner), but for its underlying historic factors (e.g., childhood experiences of being ignored). These factors habitually give rise to reasons that then motivate particular actions.

The psychological literature contains few references to causal history explanations. Locke and Pennington (1982, pp. 213, 221) hinted at the phenomenon when they considered briefly that people sometimes explain, not what reasons an agent has for acting, but why an agent has those reasons. Hirschberg (1978) provided a lucid treatment of traits as causal history factors—traits that “cause action indirectly by causing the agent's *selection* of particular beliefs and desires” (p. 60). She argued that explanations using traits as causal histories narrow down the kinds of reasons for which the action may have been performed by telling us “what kinds of reasons [the] person is likely to have” (p. 56). For example, the behavior described by “The man refused the invitation out of pride” is considered intentional, but the explanation does not explicitly mention the agent's reason for refusing. Learning that the refusal was out of pride is a causal history factor that rules out certain potential reasons (e.g., that he didn't want to spend the time) and “points to the arena in which the beliefs and desires relevant to the action are to be found” (p. 58).

Causal history of reason explanations are used in the context of explaining intentional behaviors, in contrast with cause explanations, which are used to explain unintentional behaviors. To distinguish causal history of reason explanations from cause explanations, coders must therefore infer whether the explainer took the behavior to be intentional or unintentional, and these inferences are usually reliable (see Study 2). Causal history of reason explanations and reason explanations, however, are both used for explaining intentional behaviors, so we need to explore their distinction more carefully.

In people's folk concept of intentionality, reasons have a two-faced nature. In the “forward” direction, reasons function in a reasoning process in which an agent settles on an intention by considering these reasons. In the “backward” direction, reasons qua beliefs and desires are just like other mental states that are triggered by a variety of mere causes. There is typi-

cally no intention to “have a reason” as there is an intention to act. So even though reasons play a distinguished role in generating intentional behavior via a reasoning process and an intention, reasons themselves are generated in the plain way in which unintentional behavior is generated—by “mechanical” causation. We can say that reasons can confer intentionality upon behavioral events but they are not themselves intentional.

This analysis suggests that people assume a different causal mechanism for causal history of reason explanations than they assume for reason explanations, even though they use both modes in the context of explaining intentional behaviors. The next study examines whether people indeed distinguish between these two modes if they are presented side by side as explanations of the same behaviors.

Reasons Versus Causal Histories of Reasons (Study 3)

In developmental work on the child's theory of mind it is common to differentiate implicit and explicit components of a theory of mind (e.g., Whiten, 1994). The same distinction can be applied to components of adults' folk theory of behavior. Because the scientific literature has not previously introduced the notion of causal histories of reasons and because there is no ordinary word for this mode of explanation, people probably have only implicit knowledge of the notion of causal histories. By contrast, the growing philosophical and developmental literatures on reasons and the presence of a reason concept in everyday speech suggest that people have relatively explicit knowledge of the notion of reasons. I therefore tested people's ability to discriminate between reasons and causal histories, requiring them to use only the concept of reasons explicitly.

Method

Participants and procedure. Undergraduate introductory psychology students of mixed gender and ethnicity ($N = 109$) completed a two-page measure. They were run in groups of 10-15. The sample included 9 non-native speakers, who had been learning English for 1 to 18 years ($Mdn = 12$). Exclusion of their responses did not alter the results, so we retained them. However, 20 of the 109 participants may have misunderstood the instructions, checking off only one explanation out of each set of 8 explanations. Results are reported both with and without these participants.

Material. We introduced the measure as “a test of language intuition” and requested self-ratings of language intuition at the outset. The actual measure presented three behaviors and eight explanations for each, which we had taken verbatim or with slight reformulations from another study in which subjects had explained these behaviors (O’Laughlin & Malle, 1998). Four of each behavior’s explanations were selected to be reasons (including both beliefs and desires); the other four were selected to be causal histories (including both person and situation factors; see Table 2.) This a priori classification was made on the basis of two criteria: Explanations were selected if (a) the expert coders in O’Laughlin and Malle’s (1998) study had independently identified them as clear reasons or causal history factors and (b) the present author and an assistant found them reasonably unambiguous in the absence of any context information.

Participants read each behavior description and its eight associated explanations and checked off which of the explanations they thought described one of the agent’s reasons. Reasons were introduced at the beginning of the measure as “conscious, deliberate reasons for acting that way” and defined as “something that the agent *had on his or her mind* when deciding to act (e.g., ‘Joan refused dessert because *she felt she had eaten too much already*’).” The notion of reasons was defined stringently because the term is ambiguous in common English (e.g., the “reason” why the Titanic sank [it rammed an iceberg] is a mere cause; the “reason” why Ellie is so sad [her mother died] is also a mere cause). Participants were neither told how many reasons they had to identify nor what other modes of explanations were included.

Results

Table 2 displays the eight explanations for each behavior and the rate of participants (from the whole sample) who endorsed the explanations as reasons. The endorsement rates for a priori reasons ranged from .27 to .94, with a mean of .59 and a median of .60. The endorsement rates for a priori causal histories ranged from .01 to .26, with a mean of .13 and a median of .09. The distributions are nonoverlapping and a t test on these rates is highly significant, $t(22) = 7.6, p < .001$.⁸

A more refined analysis was based on participants’ signal-detection parameters—discrimination d' and response bias $\log\beta$ —as computed from their hit rates (the rate of reason endorsements per four reasons) and

false alarm rates (the rate of reason endorsements per four causal histories) for each behavior (see McNichol, 1972, chap. 3). The d' values could range from 0 (*no discrimination*) to 6 (*perfect discrimination*). The parameter $\log\beta$ is 0 when there is no response bias, has a negative sign when there is a risky bias (accepting many false alarms to get many hits), and a positive sign when there is a conservative bias (accepting fewer hits to minimize false alarms). These parameters are displayed in Table 3 and clearly show that people had a conservative response bias (possibly because of the stringent criterion of “conscious, deliberate reasons”) but demonstrated high discrimination between reasons and causal histories.⁹ Sometimes a priori reasons were not endorsed as reasons, but a priori causal histories were rarely mistaken for reasons.

Discussion

These results support the claim that people differentiate between reasons and causal histories as explanations of intentional behavior. Even without learning about the concept of causal history explanations, people reliably discriminated them from reasons. This discrimination is more subtle than the one between reasons and causes documented in Studies 1 and 2 since the behaviors explained by reasons vs. causal histories are equated for intentionality, removing a powerful discriminative cue. Even so, people assume a different generating mechanism for causal history of reason explanations than for reason explanations. That is, they consider the way in which reasons are generated by their causal history as not involving a reasoning process or an intention, so they assume that the agent is not necessarily aware of the causal history of his reasons (and will often be unaware of it; cf. Hirschberg, 1978). Malle and O’Laughlin (1998, Studies 1-2) were able to show that people use the agent’s awareness of the facts cited in the explanation to discriminate reasons from causal histories.

The question arises how frequently and under what conditions people use causal history of reason explanations. With respect to frequency, research in our lab has found a prevalence of 6-32% across numerous data sets (Malle et al., 1998; O’Laughlin & Malle, 1998). One reliable moderator of this prevalence is the target of explanation (a collective or an individual), such that explanations of collective behaviors invite more causal history explanations than explanations of indi-

⁸After excluding the 20 participants who may have misunderstood the instructions, the endorsement rates for a priori reasons increased to .68 ($mdn = .68$) and those for a priori causal histories were .14 ($mdn = .10$), $t(22) = 8.2, p < .001$.

⁹After excluding the 20 participants who may have misunderstood the instructions, d' parameters for the three behaviors were 2.6, 2.6, and 3.2, respectively, while $\log\beta$ parameters were +0.88, +0.56, and +0.22.

Table 2. Rates of Endorsement as “Reasons” for A Priori Defined Causal History Explanations and Reason Explanations

Behavior	Rate	Causal History Explanation	Rate	Reason Explanation
By choice, Ian worked 14 hours a day last month.	.17	He is driven to achieve	.94	To make more money
	.01	That’s the cultural norm	.45	He may lose his job if he doesn’t
	.02	He is a yuppie	.65	He wants to get ahead
	.21	He works in a high-pressure corporate environment	.27	He enjoys working hard
Nancy chose not to vote in the last election.	.09	She is lazy	.77	None of the candidates appealed to her
	.08	Her mom died that week	.64	She didn’t want to support the system
	.07	She doesn’t realize that every vote counts	.34	She didn’t feel like taking the time
	.25	She’s apolitical	.56	She wasn’t interested in the issues
Brian used heavy drugs last Sunday at the party	.06	He is a junkie	.77	He was curious what it would feel like
	.05	He grew up in a drug-dealing home	.50	He’s trying to prove to a girl that he is hard-core
	.23	He is addicted to them	.64	He thought it would be cool
	.26	He gave in to peer pressure	.52	He was hoping for a spiritual experience

Table 3. Signal-Detection Analysis of People’s Distinction Between Reasons and Causal Histories

Behavior	Hit Rate		False-Alarm Rate		<i>d'</i>	<i>logβ</i>
	<i>M</i>	<i>Mdn</i>	<i>M</i>	<i>Mdn</i>		
1	0.58	0.50	0.11	0.0	2.5	+1.03
2	0.58	0.50	0.12	0.0	2.5	+0.73
3	0.61	0.75	0.15	0.0	2.7	+0.20

Note: All *d'* and *logβ* parameters are reliably different from 0, $p < .001$, except *logβ* Behavior 3.

vidual behaviors (O’Laughlin & Malle, 1998). Another moderator may be the difference between actor and observer (cf. Hirschberg, 1978), but the evidence is not yet available on this issue.

The Framework of Folk Explanations in Summary

The present theoretical framework of folk explanations of behavior specifies four modes of explanation, one for explaining unintentional behaviors and three for explaining intentional behaviors (see Figure 3). The explanation of unintentional behavior is conceptually straightforward in that the social perceiver need only identify factors that helped bring the behavior about. The sole explanatory link people must establish is that between cause(s) and behavior, and this link is expressed in cause explanations. By comparison, the explanation of intentional behavior is more complex. The folk concept of intentionality describes

intentional action as generated by the agent’s intention, which is itself based on the agent’s reasons. So the major explanatory link to be established for intentional behaviors is that between the agent’s reasons and the intention to act, and that link is expressed in reason explanations. In addition, causal history of reason explanations provide the background and origin of reasons and thus help explain the action based on them, whereas enabling factors clarify how it was possible that the agent’s intention was turned into action.

The distinction between four modes of explanation is the first contribution of the present framework. Its second contribution is a detailed model of reason explanations, a model that also identifies serious problems in attribution theory.

Part 2: Reason Explanations And The Limits of Attribution Concepts

Reason explanations capture the heart of people’s concept of intentional action: the assumption of a subjective reasoning process on whose basis the agent forms an intention to act. However, people’s reason explanations have rarely been explored in the social-psychological literature (as Fiske & Taylor, 1991, note), so no model is currently available that describes their distinct conceptual and linguistic features. The following section describes such a model, which is then contrasted with the treatment of reasons within the attribution tradition.

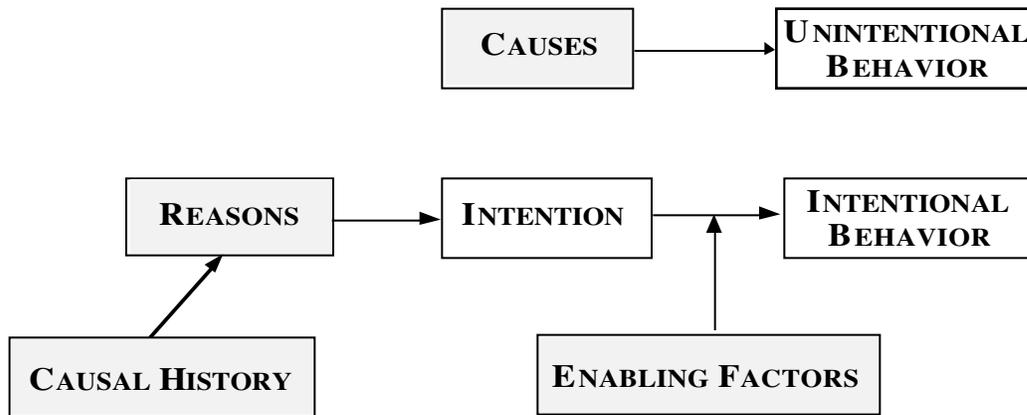


Figure 3. Four modes of folk explanation for unintentional and intentional behavior

Features of Reason Explanations

Reason explanations were defined as those behavior explanations that cite the agent’s reason(s) for intending to act or for acting intentionally, and *reasons* were defined as agents’ mental states (primarily beliefs and desires) in light of which they formed an intention to act. If these definitions capture people’s folk concepts, then people follow two important rules when providing reason explanations—a subjectivity rule and a rationality rule.

The Subjectivity Rule

Folk explainers appreciate the agent’s subjectivity by citing only those mental states as reasons that (as far as they know) the agent considered and in light of which he formed the intention to act (see Lennon, 1990, p. 18; Searle, 1983; p. 106). For example, “Why didn’t they like send you to a regular boarding school like right away?”—“Cause I think at the time they felt that what they were doing was right” [T]. To demonstrate people’s use of this subjectivity rule, Malle et al. (1998) asked people to read about actions explained by a reason (e.g., “Anne watered her plants because they were wilting.”) and then to assume that the agent was not aware of that reason (“Anne had not been aware of the fact that they were wilting”). The results showed that when the agent is described to be unaware of what appears to be a reason, social perceivers often find that the explanation does not make sense and/or infer that it must have been performed for a different reason.

The Rationality Rule

According to philosophical analyses, agents form intentions that are rationally supported by their reasons (e.g., Lennon, 1990; Mele, 1992, ch. 12). If

these analyses are correct, only those reasons should enter folk explainers’ reason explanations that rationally support the action/intention to be explained. The folk concept of *rational support* is probably not a very strict one; it merely means that, in light of the agent’s own beliefs and desires, the intended action is for the agent a reasonable thing to do. In more formal terms of philosophers since Hume, the propositional contents of reasons, if carefully spelled out, would form a practical reasoning argument that concludes with the intention to act (e.g., Harman, 1976; Snare, 1991). For example: Fred wants to have a beer, he knows there is no beer in the house, but he knows there is a corner store open at this hour, and he knows he can walk there, ... so Fred intends to walk to the store. Natural explanations usually leave many of these “premises” implicit, but the assumption is that if the explainer made them explicit, they would form a coherent practical argument.

Note that the rationality rule does not sanctify the agent’s action. The agent may still be criticized for having unacceptable desires, false or incomplete beliefs, or lack of self-control, and the action may be criticized as undesirable or morally wrong. The weak rationality assumption entailed by reason explanations merely claims that given P’s reasons, it was a reasonable thing from P’s perspective to intend to perform A. The explainer need not approve of either the reasons or the intention; she only need acknowledge that someone who has exactly P’s reasons would likely intend to perform A herself.

Even though philosophical analyses have emphasized the rationality link between reasons and intention (e.g., Bratman, 1987; Davidson, 1963; Locke & Pennington, 1982), no empirical research to date has explored this aspect of reason explanations. Several lines of inquiry seem worthwhile. One would be to show that even though any reason explanation focuses on one or two reasons, thus leaving several beliefs

and desires implicit, the explainer could be encouraged to make them explicit and thus generate a complete practical argument. Second, people should be unwilling to accept reason explanations as meaningful if they fail to provide rational support for the action (e.g., “Anne thought positive thoughts about the test because she felt she did poorly on it and she wants to do well in the class”). Moreover, they should be adept at adding assumptions that would make such explanations rational (e.g., “Anne believes that thinking positive thoughts has the power of changing past events”). Third, under rationality pressures (e.g., from an audience), people may tend to offer more reason explanations — because they are the only ones that implicate rationality — so the rate of reasons over causal histories should increase.

Linguistic Features of Reasons

As mental states, reasons have three linguistic features when used in verbal explanations: First, they usually appear as beliefs or desires (those mental states that the agent subjectively considered when forming an intention to act). An explanation citing the agent’s desire points to the desired outcome (the goal) of the action, such as when an eight-year-old explains why lovers hold hands because “*they want* to make sure that their rings don’t fall off.” An explanation citing the agent’s belief points to an actual or potential state of affairs that the agent considered when deciding to act. For example, “I just said ‘hi’ ‘cause *I knew* she wouldn’t say anything” [T].

Second, reasons can be linguistically marked with an appropriate mental state verb that indicates subjectivity (e.g., *they want*, *I knew*), or they can be left unmarked. In explaining why Anne waters her plants with Vitamin B, we may cite a desire reason that is marked (“because *she wants* them to grow faster”) or unmarked (“so they’ll grow faster”). Similarly, we may cite a belief reason that is marked (“because *she thinks* they’ll grow faster”) or unmarked (“because they’ll grow faster”).

Third, because they are subjective mental states inside the agent, reasons are always “person factors,” but they have a propositional content (*what* is believed; *what* is desired) that can refer to states inside the agent (person content) or outside the agent (situation content). In the above example, “I just said ‘hi’ ‘cause I knew she wouldn’t say anything,” the agent’s belief qua mental state is a person factor, but the content of the belief (that she wouldn’t say anything) refers to another person—a situation aspect.

Consider the following natural explanations as illustrations of the joint presence of the three linguistic features of reasons. A teenager explained to her

friend: “I didn’t say anything because Mr. Gaglean [her teacher] was there” [T]. The agent’s reason is not marked with a mental state verb, but because subjectivity and rationality still hold, the explanation must be understood as “because [I knew/realized that] Mr. Gaglean was there.” The content of this belief reason refers to a situation aspect (that Mr. Gaglean was there). By contrast, a secretary explained why a professor left a meeting: “He just went out as quickly as possible because he didn’t want to explode” [LL]. The speaker uses a mental state marker to indicate the agent’s desire to not explode. The content of this desire reason refers to the agent’s own potential behavior, a person aspect.

This linguistic analysis of reasons creates a serious problem for the traditional practice of treating reason explanations as simple “person” or “situation” attributions (a practice that works quite well for cause explanations). The problem is this: The three linguistic features of reasons span a three-dimensional feature space in which reasons can occur in eight different appearances—as a combination of at least 2 types (belief, desire), 2 contents (person, situation), and 2 forms (marked or unmarked). The attribution tradition has mapped this three-dimensional space onto a one-dimensional space of person and situation factors. Such a reductive mapping leads to a serious loss of information and may distort psychologically relevant distinctions among reasons. To assess these distortions we need to examine first how this mapping occurred.

The Reductive Mapping of Reason Explanations Onto Attribution Concepts

I consider three possible patterns of how the mapping may have occurred. The first is that all reasons may have been classified as person attributions. Some researchers argued that, because reasons are always person causes, reason explanations can only be classified as person attributions whereas cause explanations can be either person or situation attributions (e.g., Kruglanski, 1975; Miller, Smith, & Uleman, 1981; Shaver, 1985). This is a coherent position, reminiscent of Heider’s (1958) notion of personal causality, but attribution researchers did not map reasons solely onto the person attribution category. Instead, they freely applied the person-situation distinction to all explanations, thus classifying even reasons as “person attributions” or “situation attributions” (e.g., Au, 1986; McGill, 1989; Nisbett, Caputo, Legant, & Maracek, 1973; Zuckerman, 1978).

A second possibility is that researchers may have reserved the person-situation classification for the *content* of reasons only (which can indeed refer to

person aspects or situation aspects). However, researchers did not obey such content mapping either. On the contrary, traditional person–situation codings have been criticized for their oversensitivity to linguistic *surface* differences (Herzberger & Clore, 1979; Miller, Smith, & Uleman, 1981; Ross, 1977). For example, McGill (1989) coded explanations for people’s choice of college major into “situation attributions” and “person attributions.” For a situation attribution she offered the example, “finance is very challenging,” and for a person attribution she offered the example, “I like jobs that are challenging” (p. 191). The content of both reasons refers to situational aspects — finance and jobs. The major difference between the two explanations is a surface feature, namely, the presence of a mental state marker in the second example (“I like”) and its absence in the first.

This last example suggests that the mapping of reasons onto person-situation classifications may have in fact occurred via linguistic surface patterns. Ross (1977, p. 176) offered the following example of what he considered misleading surface-based attribution codings: “Jack bought the house because it was secluded” would be coded as a situation attribution, whereas “Jill bought the house because she wanted privacy” would be coded as a person attribution. Such classifications are highly sensitive to mental state markers (e.g., “she wanted”), and only if those markers are absent are the classifications sensitive to reason content (“it was secluded”). More precisely, we can formulate four hypotheses of how the mapping of reason explanations onto the person-situation dichotomy occurred.

- H1: The category of “person attributions” was assigned primarily to reasons with mental state markers because these markers include a personal pronoun referring to the agent.
- H2: Among reasons without mental state markers, person attributions were assigned primarily to those reasons that have person content (i.e., reasons whose propositional content—what the agent believes or desires—refers to aspects of the agent him- or herself).
- H3: The category of “situation attributions” was assigned primarily to reasons without mental state markers (because if there are markers, they invite assignment of person attributions).
- H4: Specifically, situation attributions were assigned primarily to unmarked reasons with situation content (i.e., reasons whose propositional content refers to aspects of the situation).

To test these hypotheses a pool of reason explanations were coded both according to the traditional

attribution scheme and according to the model described previously that identifies three distinct features of reasons.

Patterns of Reason-to-Attribution Mappings (Study 4)

Methods

Participants. Undergraduate introductory psychology students of mixed gender and ethnicity (N = 43) completed a four-page measure, some during a mass testing session, some at the end of an unrelated experiment. All received partial credit towards a course requirement for their participation.

Material. Participants completed a four-page measure that contained the same 20 behaviors as did Study 2. The behaviors were described from a first-person perspective and participants were asked to explain each of them (e.g., “Imagine you invited someone for lunch. Why would you?”). Note that even though this formulation invokes the actor perspective, it yields explanations that are not biased by privileged actor knowledge (the explainers had not actually performed the behaviors they explained). The present analysis examines only the reason explanations contained in this data set, which were generated in response to 10 different intentional behaviors (with an average intentionality rating of 5.2 on a 0-7 scale).

Attribution coding. Following the attribution tradition, all explanations were coded into three broad categories: (a) person factors, (b) situation factors, (c) interactions (explanations that involved both person and situation factors). In addition, the person category (1) was differentiated into dispositions and other person factors (e.g., temporary bodily states, past behaviors, attention). An undergraduate research assistant (who was blind to the hypotheses and not familiar with the present model of reason explanations) learned to use this fourfold scheme, and the author served as a second coder. Independent codings yielded 95% agreement on whether or not an explanation was codeable and 95% agreement ($\kappa = .90$) on distinguishing the four categories. Disagreements were resolved by discussion.

Folk explanation coding. Two researchers familiar with the present folk-conceptual framework (one of them the present author) coded all explanations for

three features of reason explanations: mental state markers (present/absent), reason content (person/situation), and reason type (belief/desire). Independent codings yielded 94% agreement on codeability, 95% agreement ($\kappa = .89$) on mental state markers, 88% agreement ($\kappa = .82$) on reason contents, and 95% agreement ($\kappa = .93$) on reason type. Disagreements were resolved by discussion.

Results

Of 515 reason explanations, 26 had to be excluded because they were deemed uncodeable in the attribution scheme. The remaining 489 reason explanations were the units of analysis in the tests of each of the mapping hypotheses.

H1: *The category of “person attributions” was assigned primarily to reasons with mental state markers.* Supporting the hypothesis, 74% of all person attributions (as compared to 7% of other attributions) were marked reasons; conversely, 97% of all marked reasons ended up in the person attribution category, whereas only 43% of unmarked reasons did (see Table 4). Examples of such marked explanations include “I asked somebody out for dinner because *I wanted* to get to know this person better” and “I applauded the musicians because *I enjoyed* their performance.” Leaving the small number (3%) of interaction attributions aside for a moment, the attributional person (vs. situation) classification was well

predicted by the presence (vs. absence) of a mental state marker ($r = .61, p < .001$).

H2: *Among reasons without mental state markers, person attributions have been assigned primarily to reasons with person content.* Among the 217 unmarked reasons, 90% of person attributions were indeed reasons with person content; conversely, 91% of unmarked reasons with person content were coded as person attributions (see Table 5). For example, “I invited somebody to lunch *to discuss* business with her”; “I drove way above the speed limit because *I was late*.”

H3: *The category of “situation attributions” was assigned primarily to reasons without mental state markers.* Indeed, 97% of situation attributions were unmarked reasons, and 53% of unmarked reasons were taken to be situation attributions, compared to only 2% of marked reasons (see Table 4). For example, “I watered my new plants because *they were dry*.” Just as hypothesis 1, hypothesis 3 is also reflected in the correlation between situation (vs. person) classification and the absence (vs. presence) of a mental state marker ($r = .61, p < .001$).

H4: *Situation attributions were assigned primarily to unmarked reasons with situation content.* Supporting this hypothesis, 90% of situation attributions were reasons with situation content, and 95% of reasons with situation content were coded as situation attributions (see Table 5). For example, “I interrupted my mother because *she was nagging me*”; “I applauded the musicians because *they were good*.”

Table 4. *Traditional Attribution Codings for Reason Explanations as a Function of Mental State Markers*

Mental State Marker	Attribution Category			Total
	Person	Situation	Interaction	
Absent (Unmarked)	93	114	10	217
Present (Marked)	263	4	5	272
Total	356	118	15	

Table 5. *Traditional Attribution Codings in Unmarked Reason Explanations as a Function of Reason Content*

Reason Content	Attribution Category			Total
	Person	Situation	Interaction	
Person aspect	84	5	3	92
Situation aspect	5	103	1	109
Interaction aspect	4	6	6	16
Totals	93	114	10	

In sum, the modal person attribution is a marked reason (74%), whereas the modal situation attribution is an unmarked reason (97%), which typically has situation content. Reflecting this clear-cut pattern, a discriminant function analysis ($N = 475$), which predicted person vs. situation attributions from the three reason features, yielded 97% correct classifications using a function with Wilks' $\Lambda = .40$ ($p < .0001$) on which mental state markers and reason content obtained significant discriminant weights.

In addition to these mapping results, relationships among reason features emerged. Of desire reasons, 72% were marked and 80% had person content. Of belief reasons, 81% were unmarked and 65% had situation content. These relationships result from syntactic regularities, which are described shortly.

Implications for Attribution Theory

Reason explanations have at least three important features (reason type, reason content, and mental state marker), but attribution researchers mapped these three features onto a simple person vs. situation distinction. I considered three possible mappings, of which only one is consistent with the past literature and the present results: Researchers' classification was based on linguistic surface patterns, such that what has been labeled "person attributions" (in the context of reason explanations) are primarily reasons with mental state markers and what has been labeled "situation attributions" are primarily reasons without mental state markers (which, in 9 out of 10 cases, have situation content). Consequently, attribution categories do not stand for pure causes. For if they did, all reasons (i.e., mental states inside the person) would be classified as person causes (Kruglanski, 1975; Miller et al., 1981), but such exclusive classification occurred neither in the past literature nor in the present study. To conclude, the practice in past research of classifying reason explanations as person vs. situation attributions did not capture people's cognitive process of assigning causes to behavior. Rather, it was based inadvertently on some features of people's expressing reason explanations. These features of expression may well have psychological significance (see below) and should be studied systematically, but they must not be mistaken for causes.

The reductive mapping of reasons onto attribution concepts invalidates several studies that coded naturally occurring reason explanations into person-situation categories. In these studies, researchers falsely interpreted their findings as indicating differences in people's cognitive processing of causes whereas they actually reflected differences in people's

expression of reasons (e.g., Nisbett, Caputo, Legant, & Maracek, 1973; Herzberger & Clore, 1979; Holtzworth-Munroe & Jacobson, 1988; McGill, 1989; Stratton, et al., 1986). By extension, the reductive mapping also invalidates studies in which people rated (on Likert scales) the degree to which an intentional behavior was "caused by aspects of the person or aspects of the situation" (e.g., Au, 1986; Miller & Ross, 1975; Regan & Totten, 1975; Storms, 1973; Zuckerman, 1978). To the extent that these studies asked people to explain intentional behaviors, the ratings are uninterpretable because people themselves had to somehow map their naturally available explanations (mostly reasons) onto the constrained person-situation scales. What might people mean by indicating a high person rating or a high situation rating? Without empirical evidence that matches up ratings with verbatim explanations, we cannot know for sure; but the range of possibilities demonstrates how ambiguous these prior results are: A high "person" rating, for example, could indicate the subject's judgment that (a) the behavior was intentional, hence based on the person's *reasons*; (b) the reason's *content* referred to a person aspect; or (c) an important person factor lay in the *causal history* of reasons. A high "situation" rating, by contrast, could indicate the subject's judgment that (a) the reason's content referred to a *situation* aspect; or (b) an important *situation* factor lay in the causal history of reasons. Many arguments have emerged over the years against using person-situation rating scales to measure people's behavior explanations (e.g., Gilbert & Malone, 1995; Malle et al., 1998); the uncertainty regarding what they tell us about people's explanations of intentional behavior is yet one more.

If we want to understand the meaning of past research that identified "person-situation" effects (e.g., actor-observer asymmetries) in the context of reason explanations (which constitute the majority of all "attributions" for intentional behavior), we need to translate these effects into patterns of reason features and then interpret these patterns. Better yet, we may want to replicate past studies and identify regularities of true reason features rather than misleading regularities of person-situation categories. Either way, we need to be able to interpret the three features of reasons.

The Interpretation of Reason Features

Because there is little research available on the nature of reason explanations, this section contains some amount of speculation. I hope, however, that the reader will consider the offered hypotheses as worthwhile paths for future research.

Mental State Markers

The presence or absence of mental state markers in reason explanations is determined by at least two processes: one grammatical, the other psychological.

The grammatical determinant of mental state markers is whether the reason is a belief or a desire, for the two reason types differ in the extent to which they require mental state verbs as markers. Desires favor mental verbs. The conjunction *because* requires mental verbs for expressing desires (e.g., “He didn’t vote because he doesn’t want to support our system of government”), whereas free-standing explanations (without *because*) can omit such markers (“They worked extra hours to get caught up on their work”). Across many data sets I have found free-standing desire expressions to be less frequent (e.g., 28% in Study 4 above), so desires more frequently come with mental state markers. By contrast, beliefs can often be expressed without a mental state marker (e.g., “She quit the job because it was uninteresting”). In Study 4, 80% of beliefs were unmarked, and across other data sets a similar rate has emerged in our lab. So whether a mental state marker is absent or present in a reason explanation is first determined by a syntactic constraint. By itself, this constraint has little psychological meaning. However, because the constraint is based on the type of reason given (desires pull for markers, beliefs do not), the presence or absence of mental state markers is a “proxy” for whatever psychological meaning accrues to the difference between desire reasons and belief reasons (see below).

The second determinant of mental state markers is the psychological function of signaling the agent’s subjectivity. That is, by using an explicit marker, as in “He quit his job because he needed a change” or “She quit the job because she thought her pay sucked,” an explainer can emphasize that the agent subjectively represented the content of the cited reason. However, this signaling function is largely restricted to belief reasons because beliefs are syntactically more fit to omit markers and because desire reasons always indicate the agent’s goal—even in their unmarked form (e.g., “They worked extra hours *to get caught up* on their work”).

The signaling of subjectivity is also more effective for observers explaining others’ behavior than for actors explaining their own behavior. When actors offer a reason explanation, their subjectivity is implied. Philosophers Moore (1993) and Wittgenstein (1953) noted that *I believe* and *I think* as markers of one’s own beliefs are always implied and can often be safely omitted. Their insight was that *asserting implies believing*. If I want to express my belief that Marion is in town, it suffices that I assert “Marion is

in town,” and everyone will infer that I believe Marion is in town. For the same reason, it makes no sense to say, “Marion is in town, but I don’t think she is,” because my first assertion implies my belief that she is in town. Because of this conversational rule that asserting implies believing, Ben can explain to Anne why he did not water the plants by simply asserting, “Because they were moist.” This assertion expresses a belief reason for it implies that Ben realized at the time that the plants were moist and for this reason did not water them.

Observers, by contrast, always have the option to emphasize or de-emphasize the actor’s subjectivity. They may emphasize it, for example, to communicate that they do not share the actor’s belief. When Oscar asks Anne why Ben did not water the plants, Anne may say, “He thought they didn’t need any water,” suggesting (with the appropriate intonation) that she did not agree with Ben’s assessment. If Anne said, however, “They didn’t need any water,” she suggests that this was Ben’s reason (he thought they didn’t need any water) *and* she agrees with it. An observer can thus embrace the actor’s reason by omitting a mental state marker or else distance herself from it by using a marker.

To summarize, because of syntactic constraints, mental state markers are more likely to be used with desires than with beliefs; markers can be used to signal subjectivity (but more suitably with beliefs than desires); observers are more likely to use markers than actors; and observers can use them to distance themselves from the agent’s subjective reasons.

Reason Type: Beliefs or Desires

Desire reasons have been regarded as strongly embedded in a culture’s shared knowledge base (Bruner, 1990) and as the primary motives of action (Searle, 1984, ch. 4). This may be because desires represent the end towards which the agent strives, whereas beliefs represent the various aspects of the path towards that end. Belief reasons typically describe specific information such as perceived circumstances, anticipated outcomes, and considered alternatives, so for an observer they may be difficult to infer. Desire reasons seem to be easier to infer, as is suggested by research on children’s theory of mind: Preschool children learn to attribute desires before they learn to attribute beliefs (e.g., Lyon, 1993; Nelson-LeGall, 1985; Premack, 1990; Wellman & Woolley, 1990; Yuill & Perner, 1988), and in their action explanations they use desires before beliefs (Bartsch & Wellman, 1989). Furthermore, most autistic children lack the ability to ascribe beliefs to other people but

have less difficulty ascribing desires to others (Baron-Cohen, 1995).

These considerations suggest that, on average, belief reasons carry more idiosyncratic information and are harder to infer. As a result, observers, who typically have less specific information available and must resort to more easily inferable explanations, should provide fewer belief reasons and more desire reasons than actors do. By the same token, explanations for unknown others should include more desire reasons than explanations for well-known others. Finally, if Bruner (1990) is correct in observing that desires are more strongly represented in a culture's shared knowledge base, then stereotypic explanations (e.g., for outgroups or prototypical group members) should more often include desire reasons than belief reasons.

Reason Content: Person or Situation

The content of reasons is determined by at least two processes: a grammatical one (based on whether the reason is a belief or desire) and a psychological one (resulting from impression management).

For grammatical reasons, person content is more often associated with desires than with beliefs. This is because most desire explanations have the syntactic structure, "I did that because I wanted *to x...*," where *x* is a state or behavior of the agent. Conversely, situation content is more often associated with beliefs than with desires, presumably because the main function of belief reasons is to represent facts of the world (the situation) that bear on the actor's fulfillment of her desires (they provide the "map by which we steer"; see Dretske, 1988). These relationships are reflected in substantial correlations between belief (vs. desire) reasons and situation (vs. person) content. In Study 4 the correlation was $r = .61$, and across other data sets I have found correlations ranging from $r = .49$ to $.71$. As a result of this substantial correlation, the determinants of desire reasons mentioned above indirectly increase the prevalence of person content, and the determinants of belief reasons indirectly increase the prevalence of situation content.

A psychological determinant of reason content lies in impression management concerns. For example, if an explainer wanted to make a to-be-explained intentional action appear particularly rational and justified, he may emphasize "objective facts," which are expressed in (unmarked) reasons with situation content (e.g., "They worked overtime because there is significant financial incentive"). In addition, when actors explain their own desirable behaviors, they may want to appear modest and not speak too much about themselves in their explanations, which would discourage reference to person content (e.g., "I look after

her because she's so absent-minded"; "I ended up carrying her a couple of times because she didn't want to walk").

To conclude, the three features of reasons have important linguistic regularities as well as psychological functions. Which of these features have been responsible for past "attribution effects" found with reason explanations can only be determined empirically. Either way, the interpretation of reason features in old and new research must take both linguistic and psychological aspects into account, and much work is needed to explore their relationship.

Integrating Attribution Concepts Into the Framework of Folk Explanations

An integration of traditional attribution concepts into the present framework must begin with the observation that people use a variety of different explanation modes, depending on the behavior's intentionality, the explainer's knowledge, and the specific question asked. Attribution theory does not distinguish between modes of explanation and thus confounds their function, use, and form. Once these modes are distinguished within the framework of folk explanation, however, we can examine to what extent attribution concepts are useful within each mode.

For cause explanations, the distinction between person and situation factors offers a meaningful classification of the kinds of causes people provide. For example, people may explain Ben's blushing during his conversation with Anne by a sexual fantasy he had (person factor) or by the giggling of Anne's girlfriends who saw the two talking (situation factor). Of course, the distinction has fuzzy boundaries because in both cases Ben's blushing was causally dependent on some situational factors (e.g., Anne's presence) and some person factors (e.g., that Ben has a crush on Anne). Whatever the true causal chain, however, the specific factors that people *mention* in cause explanations can usually be classified as either person causes or situation causes (and sometimes interactions between the two). Refinements of the person-situation dichotomy may be in order—in particular, an unconfounding of dispositional and non-dispositional person factors, as Weiner et al. (1972), Fincham and Bradbury (1992), and others have suggested by adding the stability dimension (see also F.Ex, 1998). Certain theories (e.g., Abramson, Seligman, & Teasdale, 1978; Weiner, 1986) then make predictions about the adaptability and affective consequences of these various factors cited in cause explanations.

For causal history of reason explanations, the person-situation distinction proves useful as well. Because these explanations formally refer to the *causes*

of those mental states that function as reasons, they can be classified the same way as cause explanations. Suppose that Max's reason for going to the movies on Saturday are that "Brazil" is showing and that he wants to see it again. His particular desire to see that movie may have been caused by a long-lasting fascination (person factor) for Terry Gilliam's film making, by an advertisement for the movie (situation factor), or by a conversation with his friend (person-situation interaction). The person-situation classification does not speak to the *functions* of causal history explanations (which are explored in more detail by O'Laughlin & Malle, 1998), but it may help clarify the psychological significance of citing, say, a personal as opposed to a situational causal history of a reason.

Enabling factor explanations, too, can be analyzed using the person-situation distinction. Even though there is little research explicitly studying these explanations, work on the attribution of achievement outcomes has dealt with enabling factors. Typically the agent intends to bring about such outcomes, but the intention alone does not explain a successful outcome, and enabling factor explanations—citing skill, luck, effort, etc.—bridge this explanatory gap. That is, achievement attributions are often enabling factor explanations that answer the question, "How was it possible that this outcome obtained?" For such explanations, Weiner (1986) and other researchers have demonstrated the usefulness of distinguishing between internal and external as well as stable and unstable (enabling) factors, which have distinct emotional and motivational consequences.

The distinction between person and situation causes is not useful, however, for understanding people's reason explanations of intentional behavior, which is perhaps the most important explanation mode in social contexts (Lennon, 1990; Searle, 1983; Tedeschi & Reiss, 1981). To summarize the arguments from previous sections, reasons qua mental states are always person causes, making the traditional categorization into person and situation causes inert. A modified (non-causal) person-situation distinction can be used to classify the *content* of reasons, but it does not capture other essential features of reasons (such as reason type, mental state markers), which have been confounded by attribution codings and ratings in the past.

In sum, attribution concepts have their place, if a limited one, within the framework of folk explanation. The new framework includes the person-situation distinction as one of its parameters but only for some of the explanation modes and in conjunction with a number of other parameters. As a result, the psychological study of behavior explanations is no longer limited to broad comparisons of person vs.

situation attributions but invites a multi-level analysis, distinguishing between (a) types of behavior explained (intentional vs. unintentional); (b) modes of explanation offered; and (c) essential features of variation within each mode. Such a multi-level analysis encourages numerous paths for future research, from detailed explorations of individual explanation modes to re-examinations of classic attribution topics. Below I sketch out re-examinations of two such topics: actor-observer asymmetries and self-serving biases.

Actor-Observer Asymmetries

According to the classic formulation, actors tend to attribute their own behavior to situational causes whereas observers tend to attribute the same behaviors to stable dispositions (Jones & Nisbett, 1972). This actor-observer asymmetry is expected to hold for all behaviors and all explanations (perhaps only qualified by valence, which leads to self-serving biases). However, because people use different modes of explanation for different behaviors and because these explanation modes differ in the features on which actors and observers may vary, there cannot be just one actor-observer asymmetry. The following levels of analysis must be distinguished:

On the most basic level, actors and observers can differ in the very *behaviors* they choose to explain, such as intentional vs. unintentional and observable vs. unobservable behaviors. Malle and Knobe (1997b) demonstrated that actors primarily explain unintentional and unobservable behaviors (their experiences) whereas observers primarily explain intentional and observable behaviors (others' actions).

The second level concerns the *mode of explanation* people choose. When actors and observers explain behaviors of different intentionality or disagree in the intentionality they assign to a given behavior, they will obviously differ in the mode of explanation they employ — primarily using either cause or reason explanations. But even if actors and observers agree that a behavior is intentional, they can still differ in their use of either reasons, causal histories, or enabling factors to explain that intentional behavior. Future research must address the implications of this potential asymmetry.

Third, if actors and observers explain the same type of behavior with the same mode of explanation, they can still differ in specific features of this explanation mode. When explaining unintentional behavior, actors and observers can differ in the kinds of causes they offer. To examine such differences, we may rely on the classic person-situation dichotomy—refined, perhaps, by further distinctions such as the separation of dispositional vs. non-dispositional person factors

(F.Ex, 1998; Robins, Spranca, & Mendelsohn, 1996). When people explain intentional behavior, there can be several sets of differences: When actor and observer both use causal history explanations or enabling factor explanations, we can examine whether observers more often point to (dispositional) person causes. When both use reason explanations, we can investigate actor-observer asymmetries for each reason feature. According to the interpretations of these features given in the previous section, actors should use fewer mental state markers and more belief reasons.

This multi-level approach, which awaits empirical exploration, is certainly more complex than a sole person-situation asymmetry. However, it separates effects that were confounded in the past and helps identify specific aspects of explanation on which actors and observers differ (e.g., the use of causal history factors or of mental state markers). Moreover, such specific asymmetries allow us to search for specific mechanisms that explain them, ending the unsatisfactory state of having a long list of mechanisms for one global actor-observer effect (cf. Jones & Nisbett, 1972; for a critique see Robins et al., 1996).

Self-Serving Biases

The same distinction between levels of analysis can be applied to the study of self-serving biases in behavior explanations (e.g., Bradley, 1978; Miller & Ross, 1975; Weary et al., 1982). Typically this bias has been described as people taking or denying causal responsibility for successes or failures with the goal of enhancing their private or public self-image. However, the plurality of folk explanations allows for more than one self-serving bias.

On the first level of analysis (type of behavior explained), explanations of *actions* are most likely to enhance or diminish one's self-image because they are observable (hence accessible for evaluation by others) and intentional (thus fully subject to either praise or blame). Explanations of actions should therefore be most susceptible to self-serving biases. Conversely, *experiences* are unobservable (thus difficult to evaluate by others) and unintentional (hence easier to excuse from responsibility), so explanations of experiences should be least susceptible to self-serving biases.

On the second level of analysis (mode of explanation), self-serving actors should explain their positive behaviors with reasons (implying intentionality) and their negative behaviors with causes (implying lack of intentionality) because intentionality intensifies praise and blame (Malle & Bennett, 1998; Shaver, 1985). Moreover, if such an outright distortion of the behavior's intentionality is not possible, causal history ex-

planations can be used to downplay one's reasoning process (deliberation) and focus on the origin and background of these reasons instead (O'Laughlin & Malle, 1998).

On the third level of analysis (features of each explanation mode), we should expect that citing situation causes for negative behaviors will reduce blame and citing person causes will increase blame, both when using cause explanations proper and when using enabling factor explanations to explain achievement (see Weiner, 1995). We may expect the same pattern to hold for causal history of reason explanations. In the case of reason explanations, explainers may use several of the linguistic features of reasons to fulfill self-serving goals. For example, one actor may use a marked desire reason to boast about the positive outcome he wanted to bring about, whereas another actor may use unmarked belief reasons (with situation content) to emphasize her rational consideration of facts. Future research needs to explore in detail the effects that different reason explanations have on audiences' impressions and on the choices explainers make to manage these impressions.

Limitations of the Present Framework

The empirical support base of the folk explanation framework is currently limited. Numerous predictions have not yet been tested (e.g., those surrounding rationality in reason explanations), while some are currently being tested in our lab (e.g., Malle et al., 1998; O'Laughlin & Malle, 1998). Future research, especially by other researchers, will be necessary to strengthen and refine this framework and demonstrate its usefulness in accounting for people's behavior explanations.

Another limitation is that this article focused on the conceptual and linguistic nature of folk behavior explanations; their significant social functions and consequences will have to be developed in detail elsewhere. The proposed framework should aid, however, in the study of such consequences. For example, explanations differ in their influence on an audience's impression of the agent (as already mentioned in the section on self-serving biases). Reason explanations, unlike cause explanations, can highlight the agent's rationality, making them ideal vehicles for self-presentational rationalizations. Causal history of reason explanations, though also used for intentional behaviors, minimize impressions of rationality but offer opportunities for extreme evaluation, such as by using traits ("He did it because he is evil") or distal circumstances ("I did it because I was abused as a child"). Explanations also straightforwardly affect an audience's praise or blame for the agent. Cause explanations, through their signaling of unintentionality,

keep praise or blame low, whereas reason explanations foster praise or blame via the assumptions of agency and responsibility. In situations of conflict, behavior explanations will therefore be important tools for negotiating blame and punishment. Finally, explanations also affect people's strategies to change the agent's behavior. Cause explanations invite attempts to change antecedent causal events (particularly in the environment), whereas reason explanations will be met with persuasive arguments to consider alternative reasons, trying to change the agent's subjective and (so one hopes) rational mind.

The present framework takes issue with attribution theory's assumption that all behaviors are subject to the same framework of causal explanations. This critique has significant implications, as demonstrated, for the interpretation of many classic attribution findings, but it does not necessarily affect studies that explored general features of explanations (e.g., Anderson, Krull, & Weiner, 1996), so long as these features apply equally to intentional and unintentional behaviors. Among these general features are antecedents of explanations, such as a search for meaning (e.g., Taylor, 1983), especially in response to negative or unexpected events (e.g., Bohner, Bless, Schwarz, & Strack, 1988; Weiner, 1985); consequences of explanations, such as self-interpretations of emotions (Schachter & Singer, 1962; Schwarz & Clore, 1983; Valins & Nisbett, 1972) or motivations (Lepper & Greene, 1978); and the relation between explanation and belief perseverance (Anderson, 1989; Koehler, 1991).

A deliberate limitation of the present framework is its analysis of explanations as social actions, not as cognitive processes. The model specifies the linguistic and social facets of explanations and the conceptual framework in which they are embedded; it does not specify the cognitive processes that execute explanations (e.g., Cheng & Novick, 1990; Hilton & Slugoski, 1986; Kelley, 1967; Kruglanski, 1989). This distinction between a conceptual framework and executive processes is sometimes overlooked. Compare a person who predicts rain from a druid's magical dance to a person who predicts rain from a weather chart on TV. Both may have the same neurocognitive equipment for predicting events, but the ways they see the world and act in it could not be more disparate. I have focused on how people see the world of human behavior and explain it, not what executive processes allow them to do so.

I have also not touched on the trait attributions literature because it concerns trait *inferences* from behavior rather than trait *explanations* of behavior. A perceiver draws a trait inference when she thinks that the behavior indicated the existence of a trait ("You

lent him your Power Book? You are generous"), whereas explanations occur when the perceiver thinks that the trait actually caused the behavior ("Why's she so tired?" — "She has mono"). A trait explanation implies that the perceiver made a trait inference, but a trait inference does not imply that the perceiver offered an explanation. Unfortunately, the word "attribution" has been used to refer to both of these phenomena, despite their different conceptual assumptions and social effects (Herzberger & Clore, 1979; Hilton, Smith, & Kin, 1995). The framework of folk explanations does not make predictions about (non-explanatory) trait inferences. It only locates the explanatory use of traits in cause explanations, causal history explanations, and enabling factor explanations and shows that reason explanations do not cite traits.

Finally, I should avert a common misunderstanding about the relation between folk explanations and scientific explanations of behavior. The adequacy of a social-psychological model of people's folk explanations, such as the one presented, is independent of folk psychology's ultimate validity (Fletcher, 1995; Goldman, 1993) or its potentially tautological character (cf. Smedslund, 1988). A scientific theory of folk explanations is successful if it describes, explains, and predicts people's actual explanations as based on their folk concepts of behavior (such as *intentionality* or *reason*). Whether those folk concepts are "objectively" valid as accounts of human behavior is a different question (cf. Christensen & Turner, 1993). The cognitive and behavioral sciences try to answer questions about the objective reality of the human mind and behavior, whereas a social psychology of people's behavior explanations describes people's folk theories about mind and behavior and their effects on social perception and social interaction.

Conclusions

"It is in terms of folk psychological categories that we experience ourselves and others. It is through folk psychology that people anticipate and judge one another" (Bruner, 1990, p. 15)

Heider's (1958) distinction between "personal" and "impersonal" causality defined a research agenda that was never carried out: the study of people's different modes of explanation for intentional and unintentional behavior. In the present paper I have tried to revive this research agenda by offering a theoretical framework that distinguishes between people's various modes of explanation: cause explanations for unintentional behavior and reason explanations, as well as causal history and enabling factor explanations, for intentional behavior. When attribution research applied the person-situation distinction to people's folk

explanations, it neglected their variety and the unique nature of reason explanations in particular. With the distinction between four modes of explanation and a detailed conceptual and linguistic model of reason explanations, the present framework offers a more realistic description of how people explain behavior. As a result, it opens numerous paths for future research—on novel topics such as the role of rationality and subjectivity in explanations and on classic topics such as the actor-observer asymmetry and the self-serving bias.

References

- Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology, 87*, 49-74.
- Anderson, C. A. (1989). Causal reasoning and belief perseverance. In D. Schumann (Ed.), *Proceedings of the society for consumer psychology: 1988 Annual convention* (pp. 115-120). Knoxville: University of Tennessee.
- Anderson, C. A., Krull, D. S., & Weiner, B. (1996). Explanations: Processes and consequences. In E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 271-296). New York: Guilford.
- Anscombe, G. E. M. (1957). *Intention*. Oxford, England: Basil Blackwell.
- Antaki, C. (1994). *Explaining and arguing: The social organization of accounts*. Thousand Oaks, CA: Sage.
- Antaki, C., & Leudar, I. (1992). Explaining in conversation: Towards an argument model. *European Journal of Social Psychology, 22*, 181-194.
- Au, T. K. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language, 25*, 104-122.
- Audi, R. (1993). *Action, intention, and reason*. Ithaca, NY: Cornell University Press.
- Baron-Cohen, S. (1995). *Mindblindness*. Cambridge, MA: MIT Press.
- Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60*, 946-964.
- Bohner, G., Bless, H., Schwarz, N., & Strack, F. (1988). What triggers causal attributions? The impact of valence and subjective probability. *European Journal of Social Psychology, 18*, 335-345.
- Bradley, G. W. (1978). Self-serving biases in the attribution process; A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology, 36*, 56-71.
- Brand, M. (1984). *Intending and acting*. Cambridge: MIT Press.
- Bratman, M. E. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.
- Buss, A. R. (1978). Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology, 36*, 1311-1321.
- Buss, A. R. (1979). On the relationship between reasons and causes. *Journal of Personality and Social Psychology, 37*, 1458-1461.
- Chan, D. K. (1995). Non-intentional actions. *American Philosophical Quarterly, 32*, 139-151.
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality & Social Psychology, 58*, 545-567.
- Christensen, S. M., & Turner, D. R. (Eds.). (1993). *Folk psychology and the philosophy of mind*. Hillsdale, NJ: Erlbaum Associates.
- D'Andrade, R. (1987). A folk model of the mind. In D. Holand & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 112-148). New York: Cambridge University Press.
- Davidson, D. (1963). Actions, reasons and causes. *Journal of Philosophy, 60*, 685-700.
- Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.
- Durkheim, E. (1951). *Suicide. A study in sociology* (J. A. Spaulding & G. Simpson, Trans.). Glencoe, Ill.: Free Press. (Original work published in 1897.)
- F.Ex (1998). *People's folk explanations of behavior*. Coding scheme available at <<http://www.uoregon.edu/~bfmalle/fex.html>>, University of Oregon, Eugene.
- Fincham, F. D., & Bradbury, T. N. (1992). Assessing attributions in marriage: The Relationship Attribution Measure. *Journal of Personality and Social Psychology, 62*, 457-468.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd Ed.). New York: McGraw-Hill.
- Fletcher, G. J. O. (1995). Two uses of folk psychology: Implications for psychological science. *Philosophical Psychology, 8*, 221-238.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21-38.
- Goldman, A. I. (1993). The psychology of folk psychology. *Behavioral and Brain Sciences, 16*, 15-28.
- Goldman, A. (1970). *A theory of human action*. Englewood Cliffs, NJ: Prentice-Hall.
- Harman, G. (1976). Practical reasoning. *Review of Metaphysics, 29*, 431-463.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Herzberger, S. D., & Clore, G. L. (1979). Actor and observer attributions in a multitrait-multimethod matrix. *Journal of Research in Personality, 13*, 1-15.
- Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological Bulletin, 107*, 65-81.
- 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. (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: Wiley.
- Holtzworth-Munroe, A., & Jacobson, N. S. (1988). Toward a methodology for coding spontaneous causal attributions: Preliminary results with married couples. *Journal of Social and Clinical Psychology, 7*, 101-112.
- Hume, D. (1978). *A treatise of human nature* (2nd ed., edited by L. A. Selby-Bigge & P. H. Nidditch). New York: Oxford University Press. (Original work published in 1740.)
- 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., & 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). Morristown, NJ: General Learning Press.
- Kalish, C. (1998). Reasons and causes: Children's understanding of conformity to social rules and physical laws. *Child Development, 69*, 706-720.
- 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. (1973). The process of causal attribution. *American Psychologist, 28*, 107-128.
- 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: Lawrence Erlbaum Associates, Inc.
- Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110*, 499-519.
- 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.
- Kruglanski, A. W. (1989). *Lay epistemics and human knowledge*. New York: Plenum.
- Lennon, K. (1990). *Explaining human action*. La Salle, IL: Open Court.
- Lepper, M. R., & Greene, D. (1978). *The hidden costs of reward*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Locke, D., & Pennington, D. (1982). Reasons and other causes: Their role in attribution processes. *Journal of Personality and Social Psychology, 42*, 212-223.
- Lyon, T. D. (1993). *Young children's understanding of desire and knowledge*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Malle, B. F. (1994). *Intentionality and explanation: A study in the folk theory of behavior*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Malle, B. F. (1998). Whose psychological concepts? A review of J. Smedslund's "The structure of psychological common sense." *Contemporary Psychology, 43*, 671-672.
- Malle, B. F., & Bennett, R. E. (1998). *People's praise and blame for intentions and actions. Implications of the folk concept of intentionality*. Unpublished manuscript.
- 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., & O'Laughlin, M., & Knobe, J. (1998). *The study of free-response behavior explanations: Linguistic surface vs. deep structure*. Manuscript submitted for publication.
- McClure, J. (1984). On necessity and commonsense: A discussion of central axioms in new approaches to lay explanation. *European Journal of Social Psychology, 14*, 123-149.
- McClure, J., & Hilton, D. (1997). Because you can't always get what you want: When preconditions are better explanations than goals. *British Journal of Social Psychology, 36*, 223-240.
- McGill, A. L. (1989). Context effects in judgments of causation. *Journal of Personality and Social Psychology, 57*, 189-200.
- McNicol. (1972). *A primer of signal detection theory*. London: Allen & Unwin.
- Mele, A. R. (1992). *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.
- Mele, A. R. (Ed.) (1997). *The philosophy of action*. Oxford: Oxford University Press, 1997.
- Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*, 213-225.
- Miller, F. D., Smith, E. R., & Uleman, J. (1981). Measurement and interpretation of situational and dispositional attributions. *Journal of Experimental Social Psychology, 17*, 80-95.
- Moore, G. E. (1993). Moore's paradox. In T. Baldwin (Ed.), *G. E. Moore: Selected writings* (pp. 207-212). London: Routledge.
- Nelson-LeGall, S. A. (1985). Motive-outcome matching and outcome foreseeability: Effects on attribution of intentionality and moral judgments. *Developmental Psychology, 21*, 323-337.
- Nichols, L. (1990). Reconceptualizing social accounts: An agenda for theory building and empirical research. *Current Perspectives in Social Theory, 10*, 113-144.
- 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.
- Nolen-Hoeksema, S. (1994). [Conversations between teenagers about personal problems]. Unpublished raw data.
- O'Laughlin, M., & Malle, B. F. (1998). *How people explain actions performed by groups and individuals*. Manuscript submitted for publication.
- Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36*, 1-16.
- 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.
- 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.
- Schachter, S., & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review, 69*, 379-399.
- Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1982). *Person perception* (2nd ed.). Reading, MA: Addison-Wesley.
- Schult, C. A., & Wellman, H. M. (1997). Explaining human movements and actions: children's understanding of the limits of psychological explanation. *Cognition, 62*, 291-324.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513-523.
- Scott, M. B., & Lyman, S. M. (1968). Accounts. *American Sociological Review, 33*, 46-62.
- Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Searle, J. R. (1984). *Minds, brains and science*. Cambridge, MA: Harvard University Press.
- Searle, J. R. (1995). *The construction of social reality*. New York: Free Press.
- Shaver, K. G. (1975). *An introduction to attribution processes*. Cambridge, MA: Winthrop.
- Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer.
- 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.
- Smedslund, J. (1988). *Psycho-logic*. New York: Springer.
- Snare, F. (1991). *Morals, motivation, and convention*. Cambridge: Cambridge University Press.
- Storms, M. D. (1973). Videotape and the attribution process: Reversing actors' and observers' point of view. *Journal of Personality and Social Psychology, 27*, 165-175.

- Stratton, P., Heard, D., Hanks, H. G. I., Munton, A. G., Brewin, C. R., & Davidson, C. (1986). Coding causal beliefs in natural discourse. *British Journal of Social Psychology*, *25*, 299-313.
- Svartvik, J., & Quirk, R. (Eds.) (1980). *A Corpus of English Conversation*. (Lund Studies in English, 56). Lund: Gleerups/Liber.
- Taylor, S. E. (1983). Adjustment to threatening events: A theory of cognitive adaptation. *American Psychologist*, *38*, 1161-1173.
- Tedeschi, J. T., & Reiss, M. (1981). Verbal strategies as impression management. In C. Antaki (Ed.), *The psychology of ordinary social behaviour* (pp. 271-309). London: Academic Press.
- Turnbull, W. (1986). Everyday explanation: The pragmatics of puzzle resolution. *Journal for the Theory of Social Behavior*, *16*, 141-160.
- Turnbull, W., & Slugoski, B. (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.
- Valins, S., & Nisbett, R. E. (1972). Attribution processes in the development and treatment of emotional disorders. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 63-78). Morristown, NJ: General Learning Press.
- Weary, G., Harvey, J. H., Schweiger, P., Olson, C. T., Perloff, R., & Pritchard, S. (1982). Self-presentation and the moderation of self-serving attributional biases. *Social Cognition*, *1*, 140-159.
- Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin*, *97*, 74-84.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer.
- Weiner, B. (1995). *Judgments of responsibility*. New York: The Guilford Press.
- 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., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*, 245-275.
- Wellman, H. M., Hickling, A. K., Schult, C. A. (1997). Young children's psychological, physical, and biological explanations. In H. M. Wellman, & K. Inagaki (Eds.), *The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena* (pp. 7-25). San Francisco: Jossey-Bass.
- White, P. A. (1991). Ambiguity in the internal/external distinction in causal attribution. *Journal of Experimental Social Psychology*, *27*, 259-270.
- Whiten, A. (1994). Grades of mindreading. In C. Lewis, & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 47-70). Hove, England: Erlbaum.
- Wilson, J. Q. (1997). *Moral judgment: does the abuse excuse threaten our legal system?* New York: HarperCollins.
- Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford, England: Basil Blackwell. (Original work published 1953)
- Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology*, *24*, 358-365.
- Zuckerman, M. (1978). Actions and occurrences in Kelley's cube. *Journal of Personality and Social Psychology*, *36*, 647-656.