

Conceptual Structure and Social Functions of Behavior Explanations: Beyond Person-Situation Attributions

Bertram F. Malle, Joshua Knobe, Matthew J. O’Laughlin,
Gale E. Pearce, and Sarah E. Nelson

University of Oregon

The traditional approach to studying behavior explanations involves treating them as either “person causes” or “situation causes” and assessing them via rating scales. An analysis of people’s free-response behavior explanations reveals, however, that the conceptual distinctions people use in their explanations are more complex and sophisticated than the person-situation dichotomy suggests. The authors therefore introduce a model of the conceptual structure of folk behavior explanations (the network of concepts and assumptions on which explanations are based) and test it in four studies. The modes and features of behavior explanations within this conceptual structure also have specific social functions. Two additional studies demonstrate that people alter distinct features of their explanations when pursuing particular impression management goals, and listeners make inferences about explainers’ goals and attitudes based on these features.

The study of people's behavior explanations has a strong tradition in social psychology. At least since Heider (1958), researchers have recognized the central role of explanations in people's efforts to make sense of themselves and the social world (e.g., Anderson, Krull, & Weiner, 1996; Jones, 1990; Schneider, Hastorf, & Ellsworth, 1979). Classic attribution research identified, for example, the impact of explanations on emotion (Valins & Nisbett, 1972), motivation (Weiner, 1986), relationships (Fincham & Bradbury, 1992), and moral judgment (Weiner, 1995).

The traditional approach to studying behavior explanations involves classifying explanations as either "person causes" or "situation causes." Often researchers do not directly perform these classifications but ask participants to provide ratings of how much the person and the situation contributed to the behavior in question. This approach proves problematic, however. Participants are confused about making mutually exclusive person-situation attributions (Solomon, 1978; Storms, 1973; Taylor & Fiske, 1975; Taylor & Koivumaki, 1976); researchers have failed to find the expected negative correlation between these attributions (e.g., Lepper, 1973; McArthur & Post, 1977); and different measures of person-situation attributions have not shown convergent validity (Miller, Smith, & Uleman, 1981). A more general problem of the causal-rating approach to explanations is that rating scales only weakly indicate what people actually do when they explain behavior. A naturally occurring explanation is a verbal statement (in conversation or private thought) that provides an answer to a why-question (Kidd & Amabile, 1981; Malle & Knobe, 1997b). Thus, when people are required to provide causal ratings instead of expressing verbal explanations, they are much like musicians who audition for an orchestra but are asked to analyze the music rather than invited to play.

A promising alternative to the rating-scale approach is the coding of free-response explanations (e.g., Fletcher, 1983; Islam & Hewstone, 1993; McGill, 1989; Orvis, Kelley, & Butler, 1976; Peterson, Schulman, Castellon, & Seligman, 1992). This method offers many advantages: It allows research participants to generate explanations during an experiment the way they normally would in real conversations (i.e., as answers to why-questions). It permits the use of natural texts that include previously generated explanations (e.g., spontaneous *because* statements in diaries, letters, newspapers, and books). And it encourages the application of various coding schemes derived from competing theoretical models.

Almost all work on free-response explanations, however, has adhered to the theoretical model of "person causes" and "situation causes" (e.g., McGill, 1989; Nisbett, Caputo, Legant, & Marecek, 1973). This model has stimulated an impressive amount of research on explanations but is highly restrictive. It implies that the only aspect of explanations worth studying is whether the behavior in question was caused by the person or by the situation. However, Heider (1958), who is usually credited as introducing the person-situation distinction, saw it as only one of many distinctions within the complex folk framework of behavior, and he treated it mainly as a shorthand when sketching his theory (e.g., Heider, 1958, p. 56). In the detailed exposition of his theory (pp. 79ff), Heider assigned much more importance to the distinction between unintentional and intentional causality and to the

essential role of motives or reasons in folk explanations of behavior (Malle & Ickes, in press). Similarly, recent developmental research suggests that even the 3-year-old child's theory of mind contains more sophisticated distinctions than a dichotomy between person and situation (e.g., Gopnik & Meltzoff, 1997; Wellman, 1990). For example, the major developmental progress in children's behavior explanations occurs when they conceptualize intentional action as based on beliefs and desires (e.g., Bartsch & Wellman, 1995; Kalish, 1998).

We might therefore contend that, paradoxically, the distinctions people use in their folk explanations of behavior are more complex and sophisticated than the distinctions that attribution researchers have used in their scientific descriptions of folk explanations. In this paper we will provide evidence for this contention by providing a theoretical and empirical analysis of people's free-response behavior explanations.

Linguistic Surface and Conceptual Structure of Folk Behavior Explanations

When examining free-response explanations, we must distinguish two levels of analysis: an explanation's linguistic surface and its conceptual structure. Past studies of free-response explanations have focused on linguistic surface alone. That is, coders inspected the actual words in which the explanation was formulated and classified the explanation as "person" or "situation" depending on whether the words primarily referred to the person (the agent who performed the action) or to the situation. Consider "I really hate him because *he's so strict*." When we look at the words used in this explanation, we see no mention of the actor ("I") but only of the situation ("he"), so the explanation is coded as a "situation cause." Conversely, the words in "I am glad they canceled the party 'cause *I didn't wanna go*" explicitly mention the actor, so the explanation is coded as a "person cause."

Such a surface-bound procedure quickly runs into problems. Consider "Why did he leave? — *So Erin could be by herself*." Here, the words in the explanation mention an aspect of the agent's situation, so we may be tempted to code it as a situation cause. But the explainer clearly does not refer to Erin's being by herself as a causal factor that somehow brought about the agent's departure (she was not by herself when he decided to leave). Rather, the phrase "so she could be by herself" depicts the agent's goal or purpose for leaving. Similar remarks can be made about the explanation "My father never lets us go out *because something might happen to us*." Since nothing actually has happened, the cited situation cannot itself be the cause of the father's action. Rather, the explanation refers to the actor's own beliefs: Because he *thinks* that something might happen to his daughters, he never lets them go out. These examples show that folk explanations of behavior can refer to the agent's motivation or beliefs even though the words in which they are expressed mention only the situation.

A thorough examination of people's behavior explanations must therefore consider their conceptual structure — i.e., the network of folk concepts and assumptions on which explanations are based, variously called people's *naïve psychology of action* (Heider, 1958), *theory of mind* (Premack

& Woodruff, 1978; Wellman, 1990), or *folk psychology* (Fletcher, 1995; Greenwood, 1991; Kashima, McIntyre, & Clifford, 1998). At the heart of this framework lies the folk concept of intentionality (Malle & Knobe, 1997a), which divides human behavior roughly into two classes — intentional and unintentional. This conceptual division leads people (beginning at age 3) to use distinct modes of explanation for intentional and unintentional behaviors (Buss, 1978; Kalish, 1998; Locke & Pennington, 1982; Malle, 1999; Bartsch & Wellman, 1995), and these modes of explanation differ in their conceptual structure.

Explanations for *unintentional* behavior refer to causal factors that brought about the behavior. The traditional approach to coding explanations as citing person or situation causes applies well to these explanations because they typically mention on the surface level, and also refer to on the conceptual level, the very causes presumed to have brought the behavior about.

The conceptual structure of explanations for *intentional* behavior, however, is more complex and will therefore be our sole concern in this article.¹ A first complexity is that people typically explain intentional behavior by citing the agent's reasons for acting. These *reason explanations* pose a serious problem for traditional coding approaches because they refer conceptually to subjective states of the agent (Buss, 1978; Kruglanski, 1975; Malle, 1999) but often appear on the surface as if they were situation attributions. This tension was illustrated in the two examples above — “So Erin could be by herself” and “Because something might happen to us.” In such cases, linguistic-surface coding mistakes the agent's subjective reasons for causes in the situation.

A second complexity is that people explain intentional behavior using two further modes, which we call *causal history of reason explanations* and *enabling factor explanations* (Malle, 1999). These modes of explanation, discussed in more detail below, differ from reasons in their underlying conceptual structure and social functions. Traditional attribution approaches did not distinguish reasons from these two alternative modes and therefore confounded explanations that are conceptually and functionally distinct.

In the first part of this paper (Studies 1-4) we lay out the conceptual structure of folk behavior explanations — the network of concepts and assumptions on which explanations are based — and describe the distinct modes of explanation people use to interpret intentional behavior. The second part (Studies 5-6) then illustrates some of the specific social functions served by these explanatory tools.

Study 1

We begin by distinguishing between reason explanations and causal history of reason explanations. *Reason explanations* cite the subjective reasons for which the agent acted.² That is, the explainer tries to capture the very reasons that the agent considered as grounds for acting (Audi, 1993; Kashima et al., 1998; Malle, 1999; Mele, 1992; Searle, 1983). Reasons take the form of desires, valuings, or beliefs, which are *representational mental states* (i.e., mental states that represent an object

or a proposition). When the agent forms an intention in light of certain beliefs or desires, these states constitute the reasons for which the agent forms the intention. For example, in “Why did Jarron give in? — Because he wanted to end the argument,” the explainer cites a desire in light of which the agent decided to give in. In “Why did she rush off? — She thought she was late for her class,” the explainer cites a belief in light of which the agent decided to rush off.

Causal history of reason explanations, by contrast, cite objective factors that lay in the causal history of the agent’s reasons but are not themselves reasons (cf. Hirschberg, 1978; Locke & Pennington, 1982; Malle, 1994). A causal history factor is “objective” in the sense that the explainer does not assume that the agent subjectively considered it when deciding to act — the factor caused the reason independent of the agent’s subjective awareness. For example, “Why did Jarron give in?” — “He is good-natured.” Here, Jarron wasn’t actually thinking, “I am good-natured; therefore I should give in.” In fact, he may not even have been aware that he was good-natured. Rather, the explainer presents Jarron’s good-natured character as a factor that brought about his specific reasons (e.g., his desire to end the argument). Whereas reason explanations refer to the agent’s subjective reasons for acting, causal history explanations refer to the causal background of those reasons.³

Malle (1999, Study 3) demonstrated that people distinguish reliably between reason explanations and causal history of reason explanations of the same behavior. This study, however, did not clarify how people make this distinction. Our aim in the present study is to identify a central folk-conceptual principle that discriminates between these two modes of explanation even when the linguistic surface provides no discriminatory clue.

On the linguistic surface, reasons are easily identifiable when speakers highlight the agent’s subjectivity via *mental state markers*. These are verbs that specify what type of mental state the reason is — a desire (“He went to the beach because he *wanted* to be away from everything”), a valuing (“I’m not going to move because I *like* where I live”), or a belief (“She went to the store because she *thought* they were out of rice”).

But how can we identify reasons when such mental state markers are absent? Valuing does not pose this problem because they are almost always marked — one cannot easily omit the marker in such explanations as “I like where I live” or “I enjoyed the concert.” Desires are often marked too, and their subjectivity is indicated even in unmarked desire reasons because they have characteristic grammatical forms, such as the purposive infinitive (e.g., “Then I went to the beach, just *to be away*”) and the purposive *so that* or *so* (e.g., “I’m going to stay away from Ariel *so I can spend some time with my friends*”). The challenging case is the distinction between belief reasons and causal history factors. This is because beliefs are often unmarked (Malle, 1999, Study 4) and unmarked beliefs can appear to refer to situations just as causal histories do. For example, “Why did he quit his job?” — “His boss was sabotaging him” [unmarked belief reason] vs. “His parents didn’t teach him persistence” [causal history]. If the first explanation had been expressed with a mental state marker (*he felt* that his boss was sabotaging him), no confusion would have arisen. But without mental state

markers, belief reasons might be mistaken for situation causes, as long as only the linguistic surface is considered.

What makes reasons unique is the folk-conceptual assumption that, qua reason, the explanation refers to what the agent was subjectively considering when deciding to act (Audi, 1993; Davidson, 1963; Searle, 1983). Thus, people follow a “subjectivity rule” when offering reason explanations (Malle, 1999), and this rule allows researchers to differentiate unmarked belief reasons from causal history factors in people’s explanations. If a cited explanatory factor likely operated outside subjective awareness (e.g., the worker’s upbringing), it cannot be meant as the agent’s reason for acting; if it required subjective awareness, it was likely the agent’s reason (e.g., the worker must *believe* that the boss was sabotaging him if he quit the job *because of* the sabotaging).⁴

In the present study, we examine whether people tend to judge a reason explanation as meaningless if the explanation contradicts this subjectivity rule — i.e., if the agent was not aware of the content of the unmarked belief reason. For example, “Shanna ignored her brother’s arguments because they were irrelevant (even though she was not aware that they were irrelevant).” If the explanation “they were irrelevant” implies that Shanna subjectively represented her brother’s arguments as irrelevant (and therefore ignored them), then the full statement is meaningless because the parenthetical qualification denies such a subjective representation. By contrast, causal history of reason explanations do not need to conform to the subjectivity rule, so people will often judge a causal history explanation as meaningful even if it violates this rule. For example, nothing is problematic about the statement, “Ben invited Anne for dinner because he is friendly (even though he is not aware that he is friendly).”

To test whether the subjectivity rule distinguishes between unmarked belief reasons and causal history explanations, we presented participants with behavior-explanation pairs in which the agent’s awareness of the explanation content was negated. We then asked the participants whether this explanation made sense to them or not. We expected that, with awareness negated, unmarked belief reasons should make sense less often than causal history explanations.

Method

Participants and Procedure. Undergraduate introductory psychology students ($N = 148$, 65% female) completed a one-page measure, which was part of a larger packet administered during a mass testing session. Students received partial credit towards their course requirements.

Material. The measure came in eight forms, each describing four behaviors (Anne invited Ben for dinner; Carey watered her plants; Jeremy greeted his uncle emphatically; Shaun ignored her brother’s arguments). Each behavior was paired with either an unmarked belief reason (e.g., “Anne invited Ben for dinner because *he had helped her paint her room*”) or a causal history factor (e.g., “Anne invited Ben for dinner because *she is friendly*”). A total of 8 reasons and 12 causal histories were randomly distributed across the eight forms, and each form was rated by 18-19 participants. In all behavior-explanation pairs, the agent’s awareness was negated. For example, “Anne invited Ben

for dinner because it was his last day before his move (even though she was not aware that it was his last day before his move).” After reading each pair, participants checked off whether the explanation made sense or not and indicated how intentional they thought the behavior was (on a 1-7 scale).

Intentionality judgments did not differ between reason explanations and causal history explanations.

Stimulus validation. We sampled 20 explanations (4–6 per behavior) from previous participants’ free-response explanations (Malle, 1999, Study 2) and provided a-priori classifications of them as reasons or causal history factors. To validate these classifications we asked 70 undergraduate students to inspect each explanation (paired with the corresponding action) and determine whether it could be “the agent’s conscious reason” for performing the action (cf. Malle, 1999, Study 3). These “reason rates” are displayed in Table 1. Explanations initially classified as reasons were identified as reasons by 87% of participants on average, whereas explanations initially classified as causal history factors were identified as reasons by 50% of participants on average, $t(18) = 6.8, p < .001$. Some participants showed an extreme response bias of saying “yes, this could be a reason” to almost every explanation, perhaps having misunderstood the instructions or not completing the task carefully. We therefore excluded those participants who said “yes” to more than 70% of all explanations. Thereafter, explanations initially classified as reasons were still identified as reasons by 84% of participants on average, whereas explanations initially classified as causal history factors were now identified as reasons by only 29% of participants on average, $t(18) = 7.9, p < .001$.

Visual inspection of the “reason” rates per explanation (see Table 1) suggested two distinct clusters: one encompassing 7 reasons (with rates of .78 to .98), and a second encompassing 10 causal histories (with rates of .08 to .37). The three remaining explanations fell between these clusters, with rates of .50, .57, and .65. Using generous 99.9% confidence intervals around each cluster mean, we found that none of the three intermediate explanations was part of either the causal history interval [.08; .39] or the reason interval [.72; 1.0]. The validation study was conducted after the main study, so we excluded the three intermediate explanations from further analysis.

Insert Table 1 about here

Results and Discussion

We compared each participant’s proportion of causal history explanations that were judged to make sense to that person’s proportion of reason explanations judged to make sense. (Most but not all participants judged both modes of explanation, and judgments of three explanations were excluded, so the degrees of freedom in the analyses below are smaller than the total sample size.) With awareness negated, reason explanations made sense less often ($M = .23, N = 110$) than did causal history explanations ($M = .72, N = 131$). In a repeated-measures ANOVA on those people who had judged both modes of explanation ($N = 94$), the corresponding means were .21 for reasons and .69 for causal histories, which differed significantly, $F(1,93) = 59.0, p < .001$.

These results suggest that people tolerate an agent’s unawareness of a causal history factor far more often than they tolerate an agent’s unawareness of an unmarked belief reason. Even in the

absence of the most obvious linguistic features of reasons (i.e., mental state markers), the conceptual-structure rule of an agent's subjective representation helps distinguish a reason explanation from a causal history of reason explanation. Thus, the results support the theoretical notion of a subjectivity rule in people's reason explanations and justify the coding practice of using the agent's subjective awareness as a criterion to distinguish unmarked belief reasons from causal history explanations.

However, the data also suggest that awareness as operationalized may not have perfect discriminative ability. When exposed to reason explanations with negated awareness, people did not always regard the explanation as meaningless. In 23% of the cases, they still found the explanation meaningful. Does this result imply that the subjectivity rule for reasons is not very stringently obeyed? It is possible that participants read the explanation, took it as a reason, then read the negated awareness, and concluded that the statement still makes sense because lack of awareness does not matter. Alternatively, participants may have read the explanation, taken it as a reason at first, then read the negated awareness and concluded that the agent couldn't have acted *for that reason* because the person were not aware of it. If so, then the agent must have had some other reason. In this second scenario, people actually reconstrue the reason explanation as a causal history and do so precisely because awareness is negated.

The format of Study 1 did not allow us to differentiate between these two scenarios because we did not ask participants to clarify their judgments. We therefore devised a second study in which we could ask for such clarifications. Because it may be difficult to clarify a making-sense judgment, we instead asked participants to predict whether the agent would still perform the action if he or she were not aware of the presented explanation content. Following this behavior prediction, participants should be fairly comfortable clarifying why the agent would still perform the behavior (e.g., "because unawareness doesn't matter") or why the agent would not perform the behavior (e.g., "because her reason for doing it is now gone").

Study 2

The same behavior-explanation pairs were used as in Study 1. This time participants imagined a counterfactual case in which the agent was unaware of the presented explanation content. Then they predicted whether the agent would still perform the behavior ("behavior predictions") and clarified in their own words why the agent would (or would not) perform the behavior ("clarifications"). The subjectivity rule suggests the following hypotheses for people's behavior predictions and clarifications in this counterfactual case of unawareness (see Figure 1).

Insert Figure 1 here

An action can be brought about by a causal history factor even if the agent is unaware of that factor. However, an action cannot be brought about by a reason if the agent is unaware of that reason. Thus, more people should predict that the action would still be performed if the agent's unawareness applies to a causal history factor than if it applies to a reason (Hypothesis 1).

Next, people's clarifications allowed us to examine why some people might tolerate reason explanations despite unawareness (as 23% did in Study 1). These people, predicting that the action would still be performed, should not assume that it is performed for the reason originally given. Nor should they claim that the unawareness does not matter. Rather, they should frequently offer clarifications that represent *alternative explanations*. By contrast, in light of Hypothesis 1 we expected that many people would tolerate the unawareness of a causal history explanation. These people, predicting that the action would still be performed, should spontaneously offer clarifications that represent the claim that *unawareness does not matter*, but they should feel less inclined to offer alternative explanations because the presented explanation still holds. We therefore expected that causal histories would lead to more *unawareness does not matter* clarifications than reasons would (in fact, reasons should lead to none at all); conversely, we expected that reasons would lead to more *alternative explanations* than causal histories would (Hypothesis 2).

People's clarifications might also provide direct evidence for the claim that reasons presuppose awareness and that people regard reasons of which the agent is unaware as basically inoperative. We expected that the people who (as a majority) predict that the action would not be performed when the agent is unaware of a reason content will often clarify that this is so because the *essential reason is now gone* (Hypothesis 3).

Method

Participants and Procedure. Undergraduate introductory psychology students ($N = 146$, 68% female) completed a one-page measure with 2 behavior-explanation pairs each, as part of a larger packet administered during a mass testing session. Each of the analyses below were based on somewhat fewer data points because 9 participants provided incomplete behavior predictions, 31 participants provided incomplete or uncodeable clarifications, and data for three ambiguous explanations had to be excluded (Study 2 was completed before the validation results described in Study 1 were available).

Material. The measure was described as testing "social perception skills." On the top of the page, a behavior was introduced in the form of a why-question, and an explanation was offered right below (e.g., "Why did Anne invite Ben for dinner? Because she is friendly"). After two questions about their understanding of the explanation's completeness and about the behavior's intentionality, participants made the critical behavior prediction — e.g., "If Anne had not been *aware* of the fact that she is friendly, would she still have invited Ben for dinner? (Yes, No)." Immediately following their answer, participants were asked "Why?" They clarified their prediction on a blank line. On the bottom half of the page, the same questions were asked about a second behavior and its explanation. The behavior-explanation pairs were the same as the ones analyzed in Study 1, with one exception. In addition to the unmarked belief reasons from Study 1, we added two marked belief reasons for purposes of comparison but expected no differences between the marked and the unmarked items.

Analysis. Even though each participant judged two explanations, the responses can be analyzed more elegantly if explanations are treated as units of analysis, hence as independent data. The independence assumption is supported by low and nonsignificant empirical correlations between participants' two behavior predictions ($r = .11, N = 137, p = .19$) and their two clarifications ($\chi^2 = 32.5, N = 115, df = 25, p = .14$).

Coding of clarifications. Two coders classified 231 clarifications, using the following categories: If the participant predicted that the behavior would be performed (despite unawareness), three codes were of interest: (1) Participant indicates that an *alternative reason* for performing the behavior would operate; (2) participant indicates that an *alternative causal history factor* would operate; (3) participant insists that *unawareness does not matter* — i.e., the explanatory factor can operate outside of awareness. If the participant predicted that the behavior would not be performed, two codes were of interest: (4) Participant indicates that the *essential reason is now gone*; (5) participant indicates that an *essential causal history factor is now gone*. Agreement on codeability of clarifications within this scheme was 90%, and agreement on the specific codes was 82% ($\kappa = .77$). Disagreements were settled by discussion. Eight cases were discarded because discussion did not lead to agreement.

Results

Supporting Hypothesis 1, people judged that the action would still be performed (despite unawareness) in far more cases involving causal histories (69%) than in cases involving reasons (35%). A logit analysis ($N = 243$) showed that the rate of predicted performance for causal history explanations was greater than that for both types of belief reasons ($z = 5.2, p < .001$), while the marked vs. unmarked belief reasons did not significantly differ from each other ($z = 1.08, p > .10$).

At first glance, it seems that participants violated the subjectivity rule in 35% of the cases — they predicted that the agent would perform the action even though he or she was not aware of the corresponding reason content. However, the clarifications for these cases of predicting performance suggest otherwise. Only one of 42 clarifications following reason explanations involved the claim that unawareness did not matter (2%), compared to 32 out of 81 clarifications (40%) following causal history explanations. Conversely, in 41 out of 42 clarifications (98%) following reason explanations people offered an alternative explanation, compared to 49 of 81 of clarifications (60%) following causal histories. This pattern of results supports Hypothesis 2, $\chi^2(1, N = 124) = 19.9, p < .001$.

Most of the alternative explanations for causal history factors were reasons, however. Because reason explanations are the preferred mode of explaining intentional action (Malle, 1999), these reasons may have been offered to complement, not replace, the causal history factor of which the agent was unaware. A stricter test of alternative explanations therefore examines how many “genuine substitutes” people provided in each condition — i.e., an alternative reason replacing a reason rendered unaware or an alternative causal history factor replacing a causal history rendered unaware. Presented with unaware reasons (and judging that the agent would still act), people offered 49% real

substitutes; presented with causal histories, they offered 19% real substitutes, $\chi^2(1, N = 124) = 12.5, p < .001$.

Overall — counting even the cases in which no codeable clarification was offered — the tolerance for unawareness was extremely low in the face of reason explanations: The total number of responses in which people seemed to find nothing wrong with the unawareness of reasons was only 2 out of 122. This number describes the few cases in which participants who were exposed to a reason explanation with negated awareness predicted that the behavior would still be performed and did so without offering an alternative explanation.

Finally, Hypothesis 3 stated that for reason explanations with negated awareness, the people who (as a majority) predict that the behavior would not be performed will often clarify that this is so because the essential reason for performing the behavior is now gone. Indeed, in 63 of these 71 cases (89%), participants made such a remark. Following causal history explanations, by contrast, the corresponding rate for clarifications indicating that an essential causal history factor is now gone was significantly lower at 32% (9 out of 28 cases), $\chi^2(1, N = 99) = 32.4, p < .001$.

Discussion

The present results illustrate people's reasoning about explanations of intentional behavior in light of the subjectivity rule. When people imagine that an agent is unaware of her reasons for acting, they conclude that the action either would not be performed at all (because the essential reason is now gone) or, if performed, would have to be explained differently. By contrast, when people imagine that an agent is unaware of a causal history that leads to a reason for acting, most conclude that the action would still be performed, apparently because the causal history factor does not require awareness. The results thus demonstrate that people distinguish between reasons and causal histories via the agent's subjective awareness of the explanation content: Reason explanations presuppose the agent's awareness of that content whereas causal history explanations do not.

Consequently, researchers who study folk behavior explanations are well-advised to differentiate reasons from causal histories and should do so via the same subjectivity rule that social perceivers use. That is, a given explanation of intentional behavior can only be a reason if the agent was aware of the content of that explanation, and if the agent was not aware of that content, it cannot be a reason. (For other features that distinguish reason explanations from causal history explanations see Malle, 1999, in press; O'Laughlin & Malle, 1999).

Study 3

We now introduce the third mode of explaining intentional behavior — enabling factor explanations — and show how people differentiate it from reasons and causal histories. Enabling factor explanations exist because of a particular characteristic of intentional action: the imperfect link between intention and action. An agent might have reasons to act a certain way (and plenty of causal histories for those reasons), so he or she forms an intention. But whether the action occurs may

depend on factors beyond the agent's intention and reasons — requiring, for example, skill and facilitating circumstances. A social perceiver may point to these enabling factors and thereby explain the *successful performance* of an action rather than its motives. For example, if asked “How come Phoebe got all her work done?”, a student might say, “Because she has one of these new calculators.” Phoebe's possessing a new calculator does not explain why she was trying to get her work done. Rather, given that she was trying to get it done, the calculator enabled her to succeed.

Insert Figure 2 about here

Thus, the distinction between enabling factors and the other two modes of explanation is not based on linguistic surface attributes but on their conceptual structure (see Figure 2). Reasons and causal histories explain what motivated the agent's intention to act, whereas enabling factors explain how it was possible that the agent turned the intention into a successful action (e.g., because of skill, effort, or facilitating circumstances).

Recently McClure and Hilton (1997, 1998) explored the relative use of “precondition explanations” and “motive explanations.” The latter category corresponds nicely to our notion of reason explanations, while the category of preconditions seems to include both causal histories and enabling factors — two modes of explanation that we would like to differentiate. We therefore examine the conditions under which people would most frequently offer enabling factor explanations, compared to reasons or causal histories.

First, because enabling factors explain what allowed an action to be performed, they are rarely mentioned when that performance faced no obstacles (“easy action”) but are frequently mentioned when the performance did face obstacles (“difficult action”), as suggested by McClure and Hilton (1997). In the case of a difficult action, the focus is not on explaining the intention (via reasons) but on explaining what made the performance of the action possible.

Second, because enabling factors explain why an action occurred *given* the agent's intention, they are rarely mentioned when the intention or its underlying reasons are poorly understood (“non-obvious action”); in such cases people are more concerned with explaining why the agent formed the intention in the first place. If, however, the intention and its reasons are well understood (“obvious action”), enabling factors are more frequently mentioned.

Third, because enabling factor explanations focus on how an action overcame obstacles, not on what motivated it, enabling factors should peak in response to the question “How was this possible?” (Turnbull, 1986) and should be rare in response to the question “For what reason?”

To limit the scope of our study we combined the first and second prediction into a joint test of behavior attributes (obviousness/difficulty) by comparing only the two extreme cases — difficult/obvious behaviors and easy/non-obvious behaviors. We crossed this manipulation of behavior attributes with a manipulation of question formulations, testing the third prediction above. That is, we asked participants to explain the various behaviors in response to one of four questions —

the leading questions “How was this possible?” and “For what reason?”, and two control questions, “Why?” and “How come?”

We predicted that enabling factors will peak when the behavior is difficult/obvious and when the question is “How was this possible?” We further predicted that reason explanations will peak when the behavior is easy/non-obvious and when the question is “For what reason?” Despite these systematic variations, we expected that reasons will overall be the most prevalent explanation mode for these intentional behaviors especially in response to “Why?” and “How come?”, as suggested by previous research (Graesser, Robertson, Lovelace, & Swinehart, 1980; Malle, 1999; McClure & Hilton, 1998). Finally, we expected that causal history explanations will follow a pattern similar to reason explanations but will show the lowest frequency overall because we did not include conditions that increase their prevalence (for such conditions see Malle, in press; O’Laughlin & Malle, 1999).

Methods

Participants. Introductory psychology undergraduate students ($N = 132$) completed a one-page questionnaire during a mass-testing session and received partial credit toward a course requirement.

Material. All participants were asked to explain three behaviors: “Cynthia gave her boss a check” (easy/non-obvious); “Mary, who is poor, bought a new car” (difficult/obvious; replicating a behavior used by McClure & Hilton, 1997); and “Bob finished a difficult class assignment” (difficult/obvious). Participants read one behavior description at a time and answered the explanatory question by writing on two empty lines. Question formulation (“Why?”; “How come?”; “For what reason?”; “How was this possible?”) was manipulated between subjects.

Coding. Two coders, blind to question formulation, classified more than 1300 explanations into one of three categories: enabling factor, reason, or causal history of reason. The coding instructions were simplified from a more extensive coding scheme (F.Ex, 1998), which defines the three modes of explanation according to the conceptual rules identified above. After training with about 200 explanations, a first set of 280 explanations was coded, yielding 93% agreement on codeability and 84% agreement ($\kappa = .73$) on explanation types. After discussion of disagreements, the remaining 800 explanations were coded, yielding 93% agreement on codeability and 94% agreement ($\kappa = .88$) on explanation types. Coders reached final agreement on 1276 explanations.

Analyses. Participants were assigned three scores for each behavior, one for how many reasons they mentioned, another for how many causal history factors they mentioned, and a third for how many enabling factors they mentioned. These three scores formed the three levels of the within-subject *Explanation Mode* factor. In addition, the design included a 4-level *Question Formulation* between-subjects factor and a 3-level *Behavior Type* within-subject factor. In a multivariate ANOVA we used planned contrasts for the Question factor, testing for any differences between the two control questions (“Why?” and “How come?”) and then for differences between the average of the control questions and each leading question (“How possible?” and “For what reasons?”). We also used

planned contrasts for the Behavior Type factor, testing for any differences between the two difficult/obvious behaviors and for differences between the average of these two and the easy/non-obvious behavior. In simple-effects analyses these factors were examined within each explanation mode separately. The means and analyses reported below are based on the raw scores of explanations (of each mode) per behavior. To highlight the relative use of enabling factor explanations, Table 2 displays the results in percentages (i.e., what percent of explanations per behavior fell into the reason mode, the causal history mode, and the enabling factor mode).

Results

Overall participants provided 1.4 explanations per behavior, with reasons ($M = 0.8$) dominating both enabling factors ($M = 0.5$) and causal histories ($M = 0.1$), p 's $< .001$.

We first tested the effect of behavior type (difficulty/obviousness) on explanation modes. As predicted, the two difficult/obvious behaviors elicited more enabling factors (M 's = 0.7) than did the easy/non-obvious behavior ($M = 0.1$), $t = 9.2$, $p < .001$. Conversely, the easy/non-obvious behavior elicited more reason explanations ($M = 1.1$) than did the two difficult/obvious behaviors ($M = 0.6$ and $M = 0.7$), $t = 6.0$, $p < .001$. The corresponding omnibus effects were substantial and significant for enabling factors, multivariate $F(2, 127) = 42.9$, $p < .001$, $\eta^2 = 40\%$ ⁵, and for reasons, multivariate $F(2, 127) = 19.9$, $p < .001$, $\eta^2 = 24\%$. The omnibus test for causal histories of reasons was weaker but also significant, multivariate $F(2, 127) = 9.2$, $p < .001$, $\eta^2 = 13\%$.

Insert Table 2 about here

Next we tested the effects of question formulation on explanation modes. On average, people mentioned 0.35 enabling factors in response to control questions ("Why?" and "How come?") did not differ from each other, $p > .30$). This number decreased in response to "For what reason?" ($M = 0.1$), $t = 2.5$, $p < .01$, but substantially increased in response to "How possible?" ($M = 1.2$), $t = 9.9$, $p < .001$. The opposite pattern emerged for reason explanations. On average, people mentioned 0.9 reasons in response to the control questions. This number further increased in response to "For what reason?" ($M = 1.1$), $t = 3.2$, $p < .002$, and substantially decreased in response to "How possible?" ($M = 0.4$), $t = 6.0$, $p < .001$. The corresponding omnibus tests were significant for enabling factors, $F(3, 128) = 51.1$, $p < .001$, $\eta^2 = 56\%$, and for reasons, $F(3, 128) = 25.8$, $p < .001$, $\eta^2 = 38\%$. Causal histories were not systematically affected by question formulation.

Finally, the impact of question formulation on explanation modes varied somewhat by behavior type, both for enabling factors — multivariate $F(6, 256) = 8.8$, $p < .001$, $\eta^2 = 17\%$ — and for reasons — multivariate $F(6, 256) = 7.3$, $p < .001$, $\eta^2 = 15\%$. The easy/non-obvious behavior consistently elicited a low number of enabling factors ($M = 0.0 - 0.3$) and a high number of reasons ($M = 0.9 - 1.3$), regardless of question formulation, whereas the two difficult/obvious behaviors were sensitive to question formulation. Specifically, the rate of enabling factors for these two behaviors increased in response to the "How possible?" question ($M = 1.6$), $t = 5.7$, $p < .001$, and decreased in response to the "For what reason?" question ($M = 0.1$), $t = 2.1$, $p < .05$, each compared to the rate of enabling

factors in response to control questions ($M = 0.3$). Conversely, the rate of reasons increased in response to the “For what reason?” question ($M = 1.1$), $t = 2.1$, $p < .05$, and decreased substantially in response to the “How possible?” question ($M = 0.04$), $t = 4.6$, $p < .001$, each compared to the rate of reasons in response to control questions ($M = 0.7$).

Causal history explanations were too infrequent to show reliable mean differences in the present study. However, correlations across the 12 cells of Table 2 demonstrate that the prevalence of causal histories followed a pattern similar to the prevalence of reason explanations ($r = +.21$) but dissimilar to that of enabling factor explanations ($r = -.38$). These prevalence correlations increase in size when we collapse across the two difficult/obvious behaviors and inspect the resulting 8 cells, confirming that causal histories are similar to reasons ($r = +.42$) but dissimilar to enabling factors ($r = -.53$).

Discussion

The results show, first, that common questions requesting explanations (“Why?” and “How come?”) only occasionally elicit enabling factors and the question “For what reason?” all but eliminates them. By contrast, the question “How was this possible?” maximizes enabling factors. Second, difficult/obvious behaviors elicit a higher rate of enabling factors than do easy/non-obvious behaviors. And third, this preponderance of enabling factor explanations for difficult/obvious behaviors is amplified in response to “How was this possible?” but eliminated in response to “For what reason?”

We found that reason explanations, compared to enabling factors, are dominant overall, particularly in response to common question formulations and in the context of easy/non-obvious behaviors. This inverse pattern of reasons and enabling factors is shown in their prevalence correlation of $r = -.98$ across the cells of Table 2.

These findings fit nicely within a conversational model of behavior explanations (Hilton, 1990; Turnbull, 1986; McClure & Hilton, 1998). An explanation fills either of two knowledge gaps — that for the actor’s motivation or that for the action’s successful completion. Most social behaviors are easy to perform, so only a knowledge gap for motivation opens up, which is best filled with reason explanations. If a behavior is difficult, however, a gap for the action’s completion opens up that is best filled with enabling factor explanations. Similarly, questions such as “Why?” or “For what reason?” indicate a knowledge gap for motivation and pull for reason explanations, whereas the question “How was this possible?” indicates a knowledge gap for successful completion and pulls for enabling factor explanations. On balance, social behaviors are often easily performed but hard to understand. People therefore might be more inclined to fill the gap for motivation, as reflected in the general prevalence of “Why?” questions over “How possible?” questions. Future studies may independently manipulate obviousness of motivation and difficulty of behavior to see which of the two gaps people try to fill first. But this question might be rather technical since the natural frequency of difficult behaviors that have no obvious motives may be quite low.

The distinction between enabling factors and other forms of explanation, however, is not merely technical. For one, as our data show, the two modes of explanation serve different functions in that they answer quite different questions in conversation. In addition, the distinction between the two modes may affect people's moral evaluations of the actor. A number of philosophers have suggested that reasons contribute to the action's moral worth whereas enabling factors, such as intelligence or skill, do not (Foot, 1978; Kant, 1998/1785). Consider a professor who gives especially clear lectures. The professor's behavior might be explained either with a reason (e.g., "because she wants the students to really understand") or with an enabling factor (e.g., "because she is very intelligent."). If the explainer draws attention to the professor's desire to help her students, people might be more likely to consider her morally praiseworthy, whereas if the explainer draws attention to the actor's intelligence, people may see her as talented but not necessarily morally praiseworthy. Indeed, studies have shown that people assign different levels of moral responsibility to actors based on differences in the actor's motivation, largely disregarding differences in enabling factors (see Weiner, 1995). Further research is needed to decide whether the different modes of explanation do, in fact, elicit differential moral evaluation.

We have argued that people's folk behavior explanations are more sophisticated than was suggested by classic attribution models. Studies 1 through 3 support this contention by showing that people systematically differentiate between three modes of explanation for intentional behavior — reasons, causal histories of reasons, and enabling factors. We now take a closer look at reason explanations in particular, people's primary mode of explaining intentional behavior.

The Conceptual Structure of Reason Explanations

Reasons are mental states that the agent considers when forming an intention to act. However, what the agent considers is not the mental state itself but rather the content of that state (i.e., *what* is believed, *what* is desired). When Joan decides to cancel the party because she thinks it will rain, she considers the rain, not her belief about the rain. Similarly, when Seth decides to go on vacation because he wants to relax, he (fondly) imagines the relaxation, not his desire for relaxation. We therefore need to distinguish between reasons *as mental states* (beliefs, desires, and valuings) and the *content* of those reasons (see Brentano, 1973/1872; Searle, 1983).

This conceptual distinction between state and content is also reflected in people's verbal expression of reason explanations, resulting in three features of reasons that connect conceptual assumptions to linguistic form (Malle, 1999). First, explainers either mark reasons as mental states by using a mental state verb such as "I wanted," "she thought," "he liked," or they leave reasons unmarked. We call this feature of reasons the presence or absence of *mental state markers*. Second, reasons appear as one of three major types of mental states: desire, belief, or valuing; and we call this

feature *reason type*. Third, *what* is desired, believed, or valued can be categorized as referring to the agent, the situation, or an interaction of the two. We call this feature *reason content*.

The three features of reasons are discernible only in free-response explanations (where they can be reliably coded with agreement ranging from 88% to 95%, $\kappa = .82-.93$; Malle, 1999). Thus, past studies using causal ratings not only confounded reasons with other modes of explanation (causal histories, enabling factors) but also overlooked the unique features of reasons. Even when researchers analyzed free-response explanations, they forced the variability of reason features into the narrow categories of person and situation “causes,” leading to scientific conclusions that fail to do justice to people’s conception of behavior explanations. In the study below we want to demonstrate that the traditional classification of reasons into person and situation attributions maps poorly onto the folk-conceptual structure of reasons and that these person-situation codings capture only linguistic-surface patterns that have been mistaken for causal perceptions.

Previous analyses of free-response explanations lacked a theoretical model of reasons’ conceptual structure, so researchers classified explanations via their linguistic surface. Our first hypothesis is therefore that reasons whose linguistic surface primarily mentioned the agent were coded as “person attributions.” Such reasons include those with a mental state marker (e.g., “I asked her out for dinner because *I wanted* to get to know her better”) and those that have no mental state marker but whose content refers to the agent (e.g., “I drove way above the speed limit because *I was late*”). Our second hypothesis is that reasons whose linguistic surface obviously mentioned the situation were coded as “situation attributions.” These include reasons that have no mental state marker but whose reason content refers to the situation (e.g., “I watered my new plants because *they were dry*”).

Evidence consistent with these hypotheses comes from researchers who offered prototypical examples of reason explanations that were coded into the person or situation category. Supporting our first hypothesis, prototypical examples for person attributions were marked with a mental state verb: “(I chose this major) because I like jobs that are challenging” (McGill, 1989, p. 191) or “(I chose this major) because I want to make a lot of money” (Nisbett et al., 1973, p. 158). Supporting our second hypothesis, prototypical examples for situation attributions were always unmarked reasons with situation content: “(I chose this major) because finance is very challenging” (McGill, 1989, p. 191) or “(I chose this major) because chemistry is a high-paying field” (Nisbett et al., 1973, p. 158). Note that in the last two examples, the challenge inherent in finance and the high salaries in chemistry did not directly cause the students to choose their respective majors. Rather, consistent with the subjectivity rule, the explanations refer to contents of belief reasons the same way as their more explicit surface variants do — “because [I believe] finance is very challenging” and “because [I believe] chemistry is a high-paying field.”

Further supporting evidence comes from expert coders who classified a set of over 400 reason explanations into the traditional person-situation categories as well as into the three reason features (Malle, 1999, Study 4). Of reasons classified as person attributions, 74% had mental state markers,

and of the remaining 26% without mental state markers, nine out of ten referred to agent content. Conversely, 97% of all marked reasons ended up in the person attribution category. Of reasons classified as situation attributions, 97% were expressed without mental state markers (and 90% of these referred to situation content). Conversely, 95% of reasons with situation content ended up in the situation attribution category. This clear-cut pattern is reflected in a discriminant function analysis ($N = 475$), which predicted person vs. situation attributions from reason features at a 97% classification rate (Wilks' $\lambda = .40, p < .0001$), with significant discriminant weights for mental state markers and reason content.

Apparently, past attribution analyses of explanations were highly sensitive to linguistic surface features (primarily mental state markers). But perhaps only expert coders are biased towards linguistic surface and laypeople are more sensitive to a *conceptual* person-situation structure — if there is one. To examine this possibility, we reanalyzed data collected by White (1991) in which research participants themselves provided person-situation classifications of explanations.

Study 4

White (1991) constructed descriptions of intentional and unintentional behaviors and paired these descriptions with two types of explanations: causes, paired with unintentional behaviors, and reasons, paired with intentional behaviors. White then presented 28 of such behavior-explanation pairings to participants and asked them to classify “whether the explanation referred to something internal to the actor or something external to the actor” (p. 262), leading to “internal causes,” “external causes,” “internal reasons,” and “external reasons.” Sixteen of these explanations led to relatively converging judgments (agreement among at least 13 out of 20 subjects) and these explanations, along with their external-internal judgments, were displayed in a table (White, 1991, p. 266). These tabulated external-internal judgments were the criterion that we wanted to predict.

For *intentional* behavior we expected that the features of reasons — mental state markers, reason type, and reason content — would predict whether a reason was classified by White’s participants as “external” or “internal.” In addition, the explanations of *unintentional* behavior provided a control condition. As mentioned earlier, our model treats cause explanations of unintentional behavior essentially the same as past work treated all explanations — as references to person, situation or other causes that brought about the unintentional behavior (Malle, 1999). Our classifications of causes should therefore be the same as participants’ judgments in White’s study. What we classify as person causes are White’s internal causes and what we classify as situation (and other-person) causes are White’s external causes.

Two coders, blind to the participants’ external-internal judgments, classified all explanations into categories for mode of explanation, type of cause among cause explanations, and the three features of reasons among reason explanations (F.Ex, 1998). Inter-coder reliability was perfect for all coding categories.

We matched up the participants' external/internal judgments (available for 16 of the behaviors) with our codings and found the following patterns: For cause explanations of unintentional behavior, White's internal causes matched in 4 out of 4 cases our person causes, and White's external causes matched in 4 out of 4 cases our situation causes. More important, for reason explanations White's "internal reasons" were in 4 out of 4 cases reasons with mental state markers (e.g., Anne turned on the television because *she wanted to watch a particular program*), while "external reasons" were in 4 out of 4 cases reasons without mental state markers but with situation content (e.g., George sent his parents a card because *it was their wedding anniversary*).

Thus, regardless who makes person-situation judgments of reason explanations (researchers, expert coders, or research participants), such judgments are sensitive to linguistic surface features of reasons but insensitive to their conceptual features. Specifically, the category of "person attributions" tracks the presence of mental state markers or — when markers are absent — of agent-related reason content, and the category of "situation attributions" tracks the absence of mental state markers combined with situation-related reason content.

A Puzzle Resolved

This insight into the conceptual structure of reason explanations can help resolve a well-known puzzle in past attribution research. From the perspective of classic attribution theory, this puzzle is an anomaly that cannot be explained by the theory. From the perspective of our model of folk explanations, the puzzle is merely a manifestation of the complex but predictable structure of reason explanations.

Ross (1977, p. 176) identified a problem in attribution research that has been frequently cited but never resolved (see also Monson & Snyder, 1976; Miller et al., 1981; Antaki, 1994). Ross compared two explanations: "Jack bought the house because *it was secluded*," which would be classified as a situation attribution, and "Jill bought the house because *she wanted privacy*," which would be classified as a person attribution. The puzzle for Ross was that these two attributional classifications seemed to be based on subtle differences in linguistic surface, not on fundamental differences in the underlying causal factors. This puzzle was disconcerting because attribution researchers had interpreted the person-situation dichotomy as a fundamental distinction in the causes people assign to behavior, not in the linguistic surface of explanations.

We can solve this puzzle by clarifying why it arose, namely, because reason explanations were misleadingly classified as "situation causes" or "person causes." Consider "Jack bought the house because *it was secluded*" (classified as a situation cause). It is not as though the explanation describes the seclusion as remotely acting upon Jack and causing him to buy the house. Rather, the explanation refers on the level of conceptual structure to Jack's *belief* that the house was secluded, which was his reason for buying it. (Applying the awareness test, we might ask: "If Jack had not thought that the house was secluded, would he still have bought it *because it was secluded*?" The

answer would be no.) Similarly with “I went to Spain for my holiday because it’s hot” (Antaki, 1994). The heat in Spain does not cause the agent from afar to go there; rather, the agent *believes* that it is hot in Spain, and that is her reason for going there.

Thus, the difference between the puzzling explanations cannot lie in their “locus of causality” because they are both reasons and reasons always refer (in their conceptual structure) to mental states inside the agent. As a result, the classification into person causes and situation causes is unfit to differentiate the puzzling reason explanations. What differentiates them are specific reason features. Most important, the explanations creating the puzzle differ in their use of mental state markers (e.g., “*she wanted privacy*” vs. “it was secluded”), which explains why traditional codings — sensitive to surface features such as mental state verbs — misclassified them as person and situation causes, respectively. Furthermore, the differential use of mental state markers stems in part from the fact that one explanation in the puzzling pairs is typically a desire reason whereas the other is a belief reason, and desire reasons more often carry mental state markers than do belief reasons (Malle, 1999). The puzzling pairs of explanations do not differ in their content (all four mention the situation), confirming Ross’s observation that in one respect the explanations are quite similar.

In sum, Ross’s puzzle emerged because reason explanations were forced into a single person-situation classification based on linguistic surface features. Consequently, the puzzle disappears once we recognize the folk-conceptual structure of reasons — that they are mental states that have a content — and analyze them in terms of their unique linguistic as well as conceptual features.

So far we have focused on the conceptual structure of people’s framework of behavior explanation. Now we want to highlight the specific social functions this framework serves — in impression management, the communication of attitudes, and other interpersonal endeavors. Folk explanations of behavior can be seen as a tool kit, with each explanation mode and each feature of a given mode serving as a tool for particular purposes. In Figure 3 we sketch the parts of this tool kit that are specified by our model of folk explanation. This model allows us to examine the subtle but powerful social implications of these explanatory tools. The final two studies demonstrate that when people give explanations of intentional behavior, their choice of explanatory tools reflects particular impression management goals (Study 5) and when people encounter explanations expressed by others, they rely on specific explanatory features to make inferences about the explainer’s attitudes (Study 6).

Insert Figure 3 about here

Study 5

Impression management is one of the most powerful psychological factors that alter the expression of behavior explanations (e.g., Edwards & Potter, 1993; Scott & Lyman, 1968; Tedeschi & Reiss, 1981). In the present study we examined two impression management goals — to appear rational and to appear self-centered. (Appearing self-centered is not a very common impression goal,

but we wanted to examine both a positive and a negative self-presentation.) We developed predictions about the effects of these goals on the explainer's differential use of explanation modes, types of reasons, and mental state markers.

The goal of appearing rational would be best achieved by offering reasons rather than causal history explanations because reasons highlight the agent's deliberation. Moreover, belief reasons should be favored over desire reasons because beliefs highlight the agent's thinking over mere wanting, reason over passion.

The goal of appearing self-centered would also be better achieved by reasons than by causal history of reason explanations because reasons highlight subjectivity. More important, self-centeredness would be best conveyed by reasons that carry a mental state marker, which so obviously focuses attention on the self. Compare two responses to the question "Why did you go to the party?" — "To talk with Susan" versus "Because *I wanted* to talk with Susan."

To assess the effect of these two impression management goals on behavior explanations, we asked subjects to describe six actions they performed recently. They were then prompted by an interviewer (in the presence of an audience) to explain these actions. One group was given the instruction to appear as rational as possible, another was given the instruction to appear as self-centered as possible. One control group was given no impression management instruction, and a second control group was instructed to answer the interviewer's questions as quickly as possible (to control for non-specific effects of instruction).

Method

Procedure. Eighty-nine undergraduate psychology students, who received partial credit toward a course requirement, were run in groups of 3 to 5 in a "study about how people talk in the presence of others." Participants were asked to list six actions (defined as behaviors that they chose to do and did intentionally) outside of their daily routine. Of these, three actions were to be positive and the other three neutral or negative. They were given five minutes to select and write down their six actions. Then the experimenter staged a drawing in which one participant (a confederate) was designated as the interviewer. The use of a confederate ensured that all participants were asked to explain their actions in a predetermined and consistent manner (the interviewer was blind to participants' impression conditions and to the hypotheses of the study).

One by one, participants were led into the main room where the interviewer asked them to explain four of their six actions (e.g., "Why did you go to Portland last Sunday?"). The interviewer selected those four actions that best met the criteria (i.e., being intentional and not routine), while also attempting to pick two positive and two negative actions. The why-questions and subsequent explanations were audio-recorded. After each participant answered the four why-questions, that person became an audience member who then listened to the next participant explain his or her actions. (The varying number of audience members had no effect on explanations.) After all participants had explained their actions, they were thoroughly debriefed.

Before entering the main room, participants in the experimental groups were instructed as follows: “When you meet the audience, you will be asked a few simple questions. In answering these questions, please try to appear as self-centered[rational] as possible. Afterwards, we will clarify to the rest of the group that you played this role. But for now, please keep in mind that you want to appear very self-centered [rational].” Participants in the fast group were asked to “try to answer the questions as fast as possible. Don’t think about the questions much (they will be easy); just answer them as fast as possible.” Control participants received no special instruction.

Coding. The roughly 600 answers to why-questions were transcribed and classified according to the F.Ex (1998) scheme by two coders who were blind to explainers’ impression conditions. The inter-coder agreements were as follows: 87% for the decision whether or not a given explanation was codeable; 89% ($\kappa = .67$) for explanation mode (reason, causal history, enabling factor); 96% ($\kappa = .93$) for reason type (belief, desire, valuing); 96% ($\kappa = .91$) for the presence of a mental state markers.

Results

Of the 356 behaviors people were asked to explain, 37 were excluded from analyses — 22 because they were unintentional, 3 because they were routine actions, and 12 because they were not explained. The analyses were run on the remaining 319 behaviors. These were distributed across subjects such that 100% validly explained at least two behaviors, 93% explained at least three, and 65% explained all four behaviors. Analyses were performed on each participant’s number of (a) reasons vs. causal histories, (b) beliefs, desires, and valuings, and (c) marked vs. unmarked reasons per behavior explained.

Insert Table 3

The means in Table 3 suggest that the two impression management conditions elicited, as predicted, an increase in reasons with no corresponding increase in causal histories. A multiple-measure multivariate analysis, which contained reasons and causal histories as correlated measures ($r = .23, p < .05$), confirmed that the conditions had an overall impact on the measures, multivariate $F(6, 170) = 2.27, p < .05, \eta^2 = 8\%$. However, this impact was limited to the number of reasons, which had a discriminant loading of $\beta = .92$ and a strong univariate effect, $F(3, 85) = 3.9, p = .01, \eta^2 = 12\%$. A test of individual contrasts showed that the rational group provided more reasons than the control group, $F(1, 85) = 4.8, p < .05, \eta^2 = 5\%$, and so did self-centered group, $F(1, 85) = 8.0, p < .01, \eta^2 = 9\%$. The fast group did not differ from the control group on any measure.

We predicted more specifically that people with a rational impression goal would offer more belief reasons, and the data in Table 3 support this prediction. A multiple-measure multivariate analyses confirmed that the three reason types discriminated between the rational condition and the control condition, multivariate $F(3, 83) = 4.3, p < .01, \eta^2 = 13\%$, and it was belief reasons that showed uniquely high contributions to the discriminant function ($\beta = .82$, compared to $\beta = -.47$ for desires and $\beta = .37$ for valuings) as well as a stronger univariate effect size of $\eta^2 = 9\%$ over 2% for valuings and 3% for desires. The self-centered condition also differed from the control condition, multivariate

$F(3, 83) = 3.1, p < .05, \eta^2 = 10\%$, but this effect was due to a uniform increase in all three reason types ($\eta^2 = .51$ for beliefs, $\eta^2 = .54$ for desires, and $\eta^2 = .37$ for valuing), with roughly equal univariate η^2 values of 2% to 3%. The same conclusions emerge when people's choice of reason type is treated as a three-level within-subject factor.

We also predicted that a self-centered impression goal would lead to more marked than unmarked reasons because the use of a mental state marker is a straightforward way to focus on the self. The last two columns of Table 3 show that the self-centered group indeed increased its marked reasons. Multiple-measure multivariate analyses confirmed that variations in marked and unmarked reasons discriminated between the four conditions, multivariate $F(6, 170) = 3.7, p < .01, \eta^2 = 11\%$, and people with a self-centered goal in particular offered more marked reasons than the control group, $F(1, 85) = 10.9, p < .01, \eta^2 = 11\%$. In addition, people with a rational goal offered more unmarked reasons, $F(1, 85) = 5.9, p < .02, \eta^2 = 6\%$. The fast group did not differ from the control group.

In sum, the impression management instructions (but not the speed instruction) led people to increase the number of reasons in their explanations compared to the control group. Specifically, people with a rational impression goal increased belief reasons, whereas people with a self-centered impression goal increased their marked reasons.

Insert Table 4

To better understand the scope of people's impression management strategies, exploratory analyses examined whether the increase of marked reasons under self-centered impression goals held for both marked beliefs and marked desires and whether the increase of belief reasons under rational impression goals held for both marked beliefs and unmarked beliefs. Two analysis methods were available. The first is a full factorial model testing whether each experimental group differs from the control group on reason type (belief, desire, valuing) crossed with markedness (marked, unmarked). The second is a discriminant analysis treating the six forms of reasons (marked vs. unmarked beliefs, desires, or valuing) as multiple correlated measures and examining which of them best discriminate the self-centered group from the control group and the rational group from the control group. The two analyses converged for the comparison between the self-centered and control group, suggesting that the self-centered group specifically increased marked desires to meet its impression management goal (see Table 4). In the factorial analysis, the pertinent 3-way interaction contrast was significant, multivariate $F(2, 84) = 6.0, p < .01, \eta^2 = 12\%$, relying mainly on the preponderance of marked desires in the self-centered group, $t = 2.3, p < .03$. In the discriminant analysis, marked desires contributed far more to the discriminant function ($\beta = .77$, univariate $\eta^2 = 12\%, t = 3.3, p < .01$) than the other five reason forms (β 's = .17–.36, univariate $\eta^2 = 0\%$ to 3%, *n.s.*). Thus, the self-centered group's increase in marked reasons (established earlier) is primarily driven by an increase in marked desires.

The two analyses did not converge as strongly for the comparison between the rational and the control group. In the factorial analysis, the 3-way interaction contrast fell short of significance ($p = .16$), but the discriminant analysis suggested that unmarked beliefs contribute far more to the

discriminant function ($\eta^2 = .70$, univariate $\eta^2 = 7\%$, $t = 2.6$, $p < .01$) than the other five reason forms (η^2 's = .22–.38, univariate $\eta^2 = 0\%$ to 2% , *n.s.*). Thus, we can confidently say that the rational group uses more belief reasons overall to meet its impression management goal, but whether unmarked beliefs have more impact than marked beliefs is less clear.

Discussion

Impression management goals have a discernible impact on actors' explanations of intentional actions. The two goals explored here led to a predicted increase in reason explanations relative to the control group. More important, the type of increase was specific to the impression goal. When trying to appear rational, actors offered more belief reasons, thus focusing attention on their reasoning and deliberating. When trying to appear self-centered, actors offered more marked reasons, thus focusing attention on themselves. Upon breaking down the data further we discovered that people who tried to appear self-centered primarily offered more marked *desires*, thus highlighting their wants and needs rather than their thinking. People who tried to appear rational offered far more beliefs, with a trend toward offering unmarked beliefs in particular.

It may seem surprising that people who try to appear rational increased their use of unmarked beliefs as much as (or more than) they increased their use of marked beliefs. However, while mental state markers of beliefs indicate thinking and deliberation, they can also “distance” the speaker from his own past beliefs, implying that those beliefs might not have been correct. The omission of belief markers, by contrast, can have an “embracing” effect, creating the impression that the agent simply acted on the true facts of the world. These potential distancing or embracing functions of using mental state markers were explored in a further study.

Study 6

One important function of explanations is to convey the explainer's own attitudes toward the reasons that guided the agent's action. In particular, some explanations indicate that the explainer is *embracing* the agent's reasons, portraying these reasons as sensible, desirable, or valid; other explanations indicate that the explainer is *distancing* himself from the agent's reasons, portraying these reasons as unsound, undesirable, or incorrect.

In English, for example, an explainer can distance herself from the agent's reasons by adding the phrase *for some reason*. Thus, the explanation “because *for some reason* he thought it was going to rain” not only asserts that the agent was guided by a belief that it was going to rain but also implies that this belief was unfounded. Similar remarks apply to desire reasons, such as “because *for some reason* he wanted to talk with Susan.” Not only does the explainer assert that the agent acted on a desire to talk to Susan; she also implies that this desire was foolish.

We wanted to see whether explainers can also use some of the reason features identified by our model to embrace or distance themselves from the agent's reasons. Specifically, we hypothesized that when explainers choose to give a belief reason, the unmarked form would usually be more

embracing and the marked form would usually be more distancing. Compare two responses to the question “Why is she waiting?” — the unmarked belief reason “Because Jason is still gonna come” and the marked belief reason “Because *she thinks* that Jason is still gonna come.” By using the unmarked belief, the explainer asserts that Jason actually is going to come and that the agent is waiting for that very reason. In this way, the explainer embraces the agent’s reason, implying that the agent’s belief is true. By using the marked belief, the explainer presents the agent’s belief as a subjective mental state and does not say whether Jason actually will come. Thus, the explainer distances himself from the agent’s reason, stating the belief without implying that it is true.

We also hypothesized that desire reasons would not tend to show embracing or distancing effects, in part because both marked and unmarked desires indicate the agent’s subjectivity in their linguistic surface. Again, compare two possible responses to the question “Why is she waiting?” — the unmarked desire reason “to give Jason his ticket” and the marked desire reason “because she wants to give Jason his ticket.” Here we find that neither option implies an evaluation of the agent’s desire.

To test these hypotheses, we constructed a vignette in which a speaker’s explanation could be systematically varied to present one of four reason types: a marked belief, an unmarked belief, a marked desire, or an unmarked desire. We examined the impact of these different reason types on people’s perceptions of the explainer’s attitude toward the agent.

Method

Undergraduate students ($N = 91$) read a vignette in which Cliff and Jerry are at a dinner party. Cliff asks Jerry, “Why did your girlfriend refuse dessert?” Jerry responds with one of four reason explanations (corresponding to four between-subject conditions): a marked belief (“*She thinks* she’s been gaining weight”), an unmarked belief (“She’s been gaining weight”), a marked desire (“*She wants* to lose weight”), or an unmarked desire (“To lose weight”). After reading the vignette, participants rated (on a scale from 0 to 8) how happy Jerry was with his girlfriend’s current weight and how much they liked Jerry.

Results and Discussion

As predicted, there was no significant difference between marked and unmarked desire reasons. When Jerry offered a marked desire (“She wants to lose weight”) he was not seen as significantly happier with his girlfriend’s weight ($M = 3.9$) than when he offered an unmarked desire (“To lose weight”) ($M = 3.6$), $F < 1$. Nor did participants like him significantly more when he used the marked desire ($M = 3.6$) than when he used the unmarked desire ($M = 3.2$), $F < 1$.

The difference between marked and unmarked belief reasons, in contrast, was substantial and significant. Jerry was seen as happier with his girlfriend’s weight when he used the marked belief ($M = 5.4$) than when he used the unmarked belief ($M = 2.6$), $F(1, 88) = 21.9$, $p < .01$, $\eta^2 = 20\%$. Participants also liked Jerry more when he used the marked belief ($M = 4.5$) than when he used the unmarked belief ($M = 2.9$), $F(1, 88) = 9.9$, $p < .01$, $\eta^2 = 10\%$.

Why exactly did the marked and unmarked beliefs lead to such different perceptions? There are three possibilities. First, it might be that unmarked beliefs indicate an embracing attitude (whereas marked beliefs do not). Second, it might be that marked beliefs indicate a distancing attitude (whereas unmarked beliefs do not). Third, it might be that each type of belief has a unique effect on people's perceptions, with unmarked beliefs indicating an embracing attitude and marked beliefs indicating a distancing attitude. To decide between these hypotheses we used the perceptions created by the desire reasons as a standard of comparison. Because desire reasons convey little or no information about the explainer's attitudes toward the agent, the average of the marked and unmarked desire conditions served as a neutral point against which to compare the potential embracing effect of unmarked beliefs and the potential distancing effect of marked beliefs.

If Jerry's unmarked belief ("She's been gaining weight") has an embracing effect, participants should infer that Jerry was *less happy* with his girlfriend's weight than they did in the neutral case ($M = 3.7$). Indeed, the corresponding mean was significantly lower ($M = 2.6$), $F(1, 87) = 5.0, p < .05, \eta^2 = 5\%$. If Jerry's marked belief ("She thinks she's been gaining weight") has a distancing effect, participants should infer that Jerry was *happier* with his girlfriend's weight than they did in the neutral case ($M = 3.7$). Indeed, the corresponding mean was significantly higher ($M = 5.4$), $F(1, 87) = 9.8, p < .01, \eta^2 = 9\%$. Thus, both unmarked and marked beliefs appear to affect people's perceptions, with unmarked beliefs having an embracing effect and marked beliefs having a distancing effect.

The distancing effect of "She thinks she's been gaining weight" documents nicely the social function of mental state markers. Grice (1975) argues that when a speaker adds an extra phrase to a sentence, people normally assume that the speaker has added this phrase for some particular purpose. Thus, if Jerry gives an explanation like "because *she thinks* she's been gaining weight," people assume that he must have had some purpose in adding the extra phrase "she thinks." After all, he could have simply said "because she's been gaining weight," and in accordance with the subjectivity rule people would have interpreted this utterance as the content of the agent's belief (see Studies 1 and 2). By adding the mental state marker the explainer appears to be going out of his way to avoid stating as a fact that she has been gaining weight and instead highlights the subjectivity of the agent's belief — that *she*, but perhaps no one else, thinks she has been gaining weight. Thus, people reasonably conclude that Jerry specifically used the mental state marker to distance himself from his girlfriend's belief.

General Discussion

We have argued that people's folk explanations of behavior are grounded in a more sophisticated conceptual framework than traditional attribution theories suggest. In support of this thesis, we examined the major elements in people's folk-conceptual framework of behavior and identified the various modes and features of explanation that people use to make sense of behavior.

We showed in Studies 1 through 4 that these modes and features of explanation are conceptually distinct in ways that are simply not captured by the person-situation dichotomy of attribution theory. We then showed in Studies 5 and 6 how the conceptual complexity of folk behavior explanations is used in social contexts, identifying distinct psychological functions that features of explanations perform. We now sketch some implications of the proposed theoretical model for other psychological phenomena.

Actor-Observer Asymmetries. The traditional way of studying differences in how people explain their own and others' behavior is to classify all explanations as either person or situation causes and then compare actors' and observers' use of these explanation types (following Jones & Nisbett, 1972). This approach ignores the conceptual complexity of behavior explanations. Behavior explanations cannot be reduced to a person-situation dichotomy (except on a trivial linguistic-surface level), and when they are so reduced, the interpretation of findings becomes ambiguous. Suppose a given study reports that actors gave more situation attributions whereas observers gave more person attributions. Why might that be? Perhaps actors tend to focus on unintentional behavioral events (often explained by situation causes) whereas observers focus on others' intentional actions (often explained by reasons) (Malle & Knobe, 1997b). Or the asymmetry might be a reflection of actors' tendency to avoid mental state markers. Or perhaps actors and observers differ in the content of reasons. Once we recognize the complexity of folk behavior explanations the whole question of actor-observer asymmetries must be posed anew, and detailed studies have to examine asymmetries at several levels of analysis: which behavioral events actors and observers explain in the first place (Malle & Knobe, 1997b), which explanatory modes they choose for these events, and which specific features of explanations they favor (Malle, 1999).

Blame Regulation. People alter explanations to regulate blame. For example, a person who has just committed a crime will try to avoid blame when explaining his action. He might tend to choose causal history explanations because they turn attention away from his choice and reasoning, and causal histories referring to the agent's situation would seem the most effective ("I did it because my parents would always...") (Wilson, 1997). By contrast, an explainer in the observer perspective who is trying to maximize the blame attributed to the agent might use primarily marked reasons ("He did it because he specifically wanted to destroy her") or causal histories that refer to the agent's undesirable dispositions ("He did it because he is lazy, irresponsible, and cruel"). Research into blame regulation would shed new light on past studies of self-serving biases, which also reduced people's explanations to a simple person-situation dichotomy.

Cross-cultural differences. The study of cross-cultural differences in explanation has also been guided by the person-situation dichotomy (Menon, Morris, Chiu, & Hong, 1999; Miller, 1987; Morris & Peng, 1994). But differential cultural or national patterns of behavior explanation are difficult to interpret if we fail to distinguish between the multiple levels on which explanations can be analyzed. Cultures can differ in the behavioral events they routinely explain, their judgments of

intentionality for these behaviors, the selected modes of explanation, and the specific features of these explanations. And in order to study these potential psychological differences researchers must also take into account regularities in language that can produce differences in the linguistic surface of explanations. In English, almost all valuing is marked with mental state verbs, highlighting the agent, whereas in French, for example, valuing may highlight the object valued, as in “parce que ça plait.” To avoid mistaken conclusions drawn from mere linguistic differences, researchers must use conceptual tests such as agent-subjectivity and rationality to complement their analyses of people’s expressed explanations.

Conclusion. Naturally occurring behavior explanations can be studied successfully as long as they are classified into categories based on people’s own folk theory of behavior. Thus, if research is to go beyond mere surface differences in explanations and reach conclusions about what these differences really mean, we must turn from rough person-situation classifications to a careful consideration of the sophisticated concepts and rules that underlie people’s behavior explanations. Studies must keep track of the behaviors people explain (intentional, unintentional), their selected modes of explanation (causes, reasons, causal histories, enabling factors), and the specific features within each mode (e.g., mental state markers, reason type). The present studies suggest that these theoretical distinctions validly describe the conceptual structure of people’s folk behavior explanations and encourage the study of explanations in their complex social context.

Footnotes

- ¹ Malle (1999) provides a more detailed account of cause explanations of unintentional behavior and contrasts them with the remaining modes of explanation.
- ² The term *reason* is commonly used in two ways: (a) the way we define it — as *the agent's* reason for acting intentionally (e.g., “My reason for doing that was . . .”), and (b) as a synonym for “cause” (e.g., “The reason why the Titanic sank was . . .”). The two terms are homophones but clearly distinct — the synonym for cause cannot be *the agent's* reason (e.g., we cannot say, “The Titanic's reason for sinking was . . .”). Our theoretical claims only apply to the first meaning.
- ³ Note that causal history of reason explanations also differ from *cause explanations* in that causal histories apply to intentional behavior whereas cause explanations apply to unintentional behavior. The causal structure of causal history explanations is identical to that of cause explanations (both describe a mechanical generation of events), but causal history factors are presumed to generate the reasons for an agent's intention and intentional action.
- ⁴ We have no data on the precise level of subjective awareness folk explainers assume for an agent's consideration of reasons. It seems clear that a wholly unconscious desire or belief does not count as a reason *for which* the agent acted. But a reason need not be considered with full deliberation either. What seems to be required is at least a dim level of awareness, such that, if asked, a truthful agent would endorse that desire or belief as his or her reason for acting.
- ⁵ When describing multivariate effects in this article, we report a multivariate η^2 , which corresponds to V/s — i.e., Pillai's criterion divided by the number of discriminant functions in the model.

References

- 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 Press.
- Antaki, C. (1994). *Explaining and arguing: The social organization of accounts*. Thousand Oaks, Calif.: Sage Publications.
- Audi, R. (1993). *Action, intention, and reason*. Ithaca, NY: Cornell University Press.
- Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.
- Brentano, F. C. (1973). *Psychology from an empirical standpoint*. (Transl. from the German by A. C. Rancurello, D. B. Terrell, & L. L. McAlister.) New York: Humanities Press. (Originally published in 1872.)
- Buss, A. R. (1978). Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology*, 36, 1311-1321.
- Davidson, D. (1963). Actions, reasons and causes. *Journal of Philosophy*, 60, 685-700.
- Edwards, D., & Potter, J. (1993). Language and causation: A discursive action model of description and attribution. *Psychological Review*, 100, 23-41.
- F.Ex (1998). *F.Ex: Coding scheme for people's folk explanations of behavior*. <<http://darkwing.uoregon.edu/~interact/fex.html>>, University of Oregon.
- Fincham, F. D., & Bradbury, T. N. (1992). Assessing attributions in marriage: The relationship attribution measure. *Journal of Personality and Social Psychology*, 62, 457-468.
- Fletcher, G. J. O. (1995). *The scientific credibility of folk psychology*. Mahwah, NJ: Erlbaum.
- Fletcher, G. J. O. (1983). The analysis of verbal explanations for marital separation: Implications for attribution theory. *Journal of Applied Social Psychology*, 13, 245-258.
- Foot, P. (1978). Virtues and vices. In P. Foot, *Virtues and vices and other essays in moral philosophy*. Berkeley: University of California Press.

Graesser, A. C., Robertson, S. P., Lovelace, E. R., & Swinehart, D. M. (1980). Answers to why-questions expose the organization of story plot and predict recall of actions. *Journal of Verbal Learning and Verbal Behavior*, 9, 110-119.

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

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

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

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

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

Islam, M. R., & Hewstone, M. (1993). Intergroup attributions and affective consequences in majority and minority groups. *Journal of Personality and Social Psychology*, 64, 936-950.

Jones, E. E. (1990). *Interpersonal perception*. New York: W.H. Freeman.

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.

Kant, I. (1998). *Groundwork of the metaphysics of morals*. (Transl. from the German by Mary Gregor.) Cambridge: Cambridge University Press. (Originally published 1785.)

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

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

- Kruglanski, A. H. (1975). The endogenous-exogenous partition in attribution theory. *Psychological Review*, 82, 387-406.
- Lepper, M. R. (1973). Dissonance, self-perception, and honesty in children. *Journal of Personality and Social Psychology*, 25, 65-74.
- Locke, D., & Pennington, D. (1982). Reasons and other causes: Their role in attribution processes. *Journal of Personality and Social Psychology*, 42, 212-223.
- Malle, B. F. (1994). *Intentionality and explanation: A study in the folk theory of behavior*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, 3, 21-43.
- Malle, B. F. (in press). Folk explanations of intentional action. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and Intentionality: Foundations of Social Cognition*. Cambridge, MA: MIT Press.
- Malle, B. F., & Ickes, W. (in press). Fritz Heider: Philosopher and psychologist. Chapter to appear in G. A. Kimble & M. Wertheimer (Eds.), *Portraits of Pioneers in Psychology* (Vol. 4). Washington, DC: American Psychological Association.
- Malle, B. F., & Knobe, J. (1997a). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101-121.
- Malle, B. F., & Knobe, J. (1997b). Which behaviors do people explain? A basic actor-observer asymmetry. *Journal of Personality and Social Psychology*, 72, 288-304.
- McArthur, L. Z., & Post, D. L. (1977). Figural emphasis and person perception. *Journal of Experimental Social Psychology*, 13, 520-535.
- McClure, J., & Hilton, D. (1997). For you can't always get what you want: When preconditions are better explanations than goals. *British Journal of Social Psychology*, 36, 223-240.
- McClure, J., & Hilton, D. (1998). Are goals or preconditions better explanations? It depends on the question. *European Journal of Social Psychology*, 28, 897-911.
- McGill, A. L. (1989). Context effects in judgments of causation. *Journal of Personality and Social Psychology*, 57, 189-200.

- Mele, A. R. (1992). *Springs of action: Understanding intentional behavior*. New York : Oxford University Press.
- Menon, T., Morris, M. W., Chiu, C.-Y., & Hong, Y.-Y. (1999). Culture and the construal of agency: Attribution to individual versus group dispositions. *Journal of Personality & Social Psychology*, 76, 701-717.
- 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.
- Monson, T. C., & Snyder, M. (1976). Actors, observers, and the attribution process: Toward a reconceptualization. *Journal of Experimental Social Psychology*, 13, 89-111.
- Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality & Social Psychology*, 67, 949-971.
- Nisbett, R. E., Caputo, C., Legant, P., & Marecek, J. (1973). Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology*, 27, 154-164.
- O’Laughlin, M., & Malle, B. F. (1999). *How people explain actions performed by groups and individuals*. Unpublished Manuscript, University of Oregon.
- Orvis, B. R., Kelley, H. H., & Butler, D. (1976). Attributional conflict in young couples. In J. H. Harvey, W. Ickes, & R. Kidd, *New directions in attribution research* (Vol. 1, pp. 353-386). Hillsdale, NJ: Erlbaum.
- Peterson, C., Schulman, P., Castellon, C., & Seligman, M. E. P. (1992). The explanatory style scoring manual. In C. P. Smith, & J. W. Atkinson (Eds.), *Motivation and personality: Handbook of thematic content analysis* (pp. 376-382). New York: Cambridge University Press.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515-526.
- 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.
- Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1982). *Person perception* (2nd ed.). Reading, MA: Addison-Wesley.

- 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.
- Solomon, S. (1978). Measuring dispositional and situational attributions. *Personality and Social Psychology Bulletin*, 4, 589-594.
- 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.
- Taylor, S. E., & Fiske, S. T. (1975). Point-of-view and perceptions of causality. *Journal of Personality and Social Psychology*, 32, 439-445.
- Taylor, S. E., & Koivumaki, J. H. (1976). The perception of self and others: Acquaintanceship, affect, and actor-observer differences. *Journal of Personality and Social Psychology*, 33, 403-408.
- Tedeschi, J. T., & Riess, 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.
- 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.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer.
- Weiner, B. (1995). *Judgments of responsibility*. New York: The Guilford Press.
- Wellman, H. (1990). *Children's theories of mind*. Cambridge, MA: MIT Press.
- White, P. A. (1991). Ambiguity in the internal/external distinction in causal attribution. *Journal of Experimental Social Psychology*, 27, 259-270.
- Wilson, J. Q. (1997). *Moral judgment: Does the abuse excuse threaten our legal system?* New York: HarperCollins.

Author Note

Bertram F. Malle, Joshua Knobe, Matthew J. O’Laughlin, Gale E. Pearce, Sarah E. Nelson,
Department of Psychology, University of Oregon.

This research was supported by an NSF CAREER award (SBR-9703315) to the first author
and an NSF graduate fellowship to the third author.

We thank Craig Anderson and John McClure for their comments on an earlier draft.

Please address correspondence regarding this article to Bertram F. Malle, Department of
Psychology, 1227 University of Oregon, Eugene, OR 97403, or send Email to
bfmalle@darkwing.uoregon.edu.

Table 1.

Participants' classifications of explanations used in Studies 1 and 2 as reasons or causal histories

	Rate of being judged a reason		Final
	All	Bias < .70	Classification
Jeremy greeted his uncle emphatically ...			
“because he always does”	.36	.08	CHR
“because he is friendly”	.45	.22	CHR
“because he was taught to”	.46	.28	CHR
“because he was in a good mood”	.51	.28	CHR
“because he thought he might get some money if he did”	.81	.83 ^b	REA
“because his uncle looked depressed”	.91	.87	REA
Carey watered her plants ...			
“because she takes good care of plants”	.47	.20	CHR
“because she was at home”	.43	.24	CHR
“because she does it every week”	.59	.37	CHR
“because they needed it”	.89	.83	REA
“because the leaves were wilting”	.97	.95	REA
Anne invited Ben for dinner ...			
“because she is friendly”	.43	.20	CHR
“because they are friends”	.61	.39	CHR
“because it was her roommate’s birthday”	.81	.78	REA
“because he had helped her paint her room”	.87	.83	REA
“because she hadn’t talked to him all week”	.87	.88	REA
“because it was his last day before his move”	.99	.98	REA
Shaun ignored her brother’s arguments ...			
“because she is stubborn”	.29	.08	CHR
“because she found them irrelevant”	.88	.85 ^b	REA
“because she was on her way out”	.57	.50 ^a	-
“because she was angry at him”	.78	.65 ^a	-
“because they were incorrect”	.65	.57 ^a	-

Note: REA = reason explanation; CHR = causal history of reason explanation.

^a Reason rates that did not fall either in the CHR cluster or the REA cluster and were excluded.

^b Marked beliefs were included in Study 2 only