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

1

Attributions as Behaviour Explanations: Towards a New Theory

Bertram F. Malle, University of Oregon

Attribution theory is a hallmark of social-psychological thinking. Thousands of articles have been published in over forty years of research, and textbooks and handbooks of social psychology typically devote a chapter or a large section to attribution phenomena. This body of research can be usefully divided into a *general attributional approach* to social-psychological phenomena and *theories of specific attribution phenomena*, such as Kelley's (1967) theory of explanation or Jones and Davis's (1965) theory of dispositional inference. The general attributional approach recognizes that humans try to make sense of themselves and their surroundings and that this sense-making activity (explanations, finding meaning, creating stories) is an integral part of the social phenomena under investigation. This approach has made countless contributions to the literature, shedding light on achievement motivation, responsibility judgements, helplessness, sleep disturbance, obesity, depression, emotion and well-being research (e.g., Abramson, Seligman, & Teasdale, 1978; Jones, Kanouse, Kelley, Nisbett, Valins, & Weiner, 1972; Schwarz & Clore, 1983; Weiner, 1995).

Attribution theories, by contrast, are theories of the phenomenon of attribution itself. Unfortunately, at least two distinct phenomena have been referred to using the term *attribution* (Hamilton, 1998; Hilton, Smith, & Kin, 1995; Malle, 2004). According to one common meaning, forming an attribution is *giving an explanation* (especially of behaviour); according to another, forming an attribution is *making a dispositional (trait) inference* from behaviour. Even though explanations and trait inferences are occasionally related, they are distinct in many ways. For example, explanations sometimes refer to traits but often do not; trait inferences can be explanatory but usually are not; traits can be inferred from any behaviour, whereas explanations are triggered only by surprising or confusing behaviour; explanations are answers to "why" questions, and trait inferences are not.

My focus in this chapter is on the phenomenon of behaviour explanations.[1] In particular, I discuss a new theory of explanation that provides an alternative to the traditional attribution theory that dominates the textbooks and handbooks, which is typically a version of Kelley's (1967) model of attribution as covariation detection. I begin with a brief critique of the traditional theory and then, out of this critique, develop a list of requirements that an improved theory has to meet. I then introduce the new theory, report supporting empirical data and apply it to a number of psychological phenomena. I conclude with an assessment of how much progress we have made in understanding behaviour explanations and what has yet to be learned.

Traditional Attribution Theory

Fiedler, Walther, & Nickel, 1999 NOT IN REFS

Kelley's (1967) original theory, as well many others after it, made two core claims (Cheng & Novick, 1990; Fiedler, Walther, & Nickel, 1999; Försterling, 1989; Hewstone & Jaspars, 1987):

1. Claim PS: The causal concepts on which people rely when forming behaviour explanations consist of a dichotomy of internal versus external, or *person versus situation*, causes.
2. Claim COV: The cognitive process that underlies explanations is *covariation analysis*.

When we examine each of these core claims of attribution theory in turn, we see that there is shockingly little support for either. First, what is the empirical evidence for claim PS? In most attribution studies the truth of PS was assumed, not tested. Participants had to fill out scales for "person/disposition" and "situation" causes (e.g., Storms, 1973) or researchers classified free-response explanations into person and situation categories (e.g., McGill, 1989; Nisbett, Caputo, Legant, & Marecek, 1973). In neither case was assumption PS falsifiable. The few attribution studies that did not automatically make the person/situation assumption found several other dimensions of explanation (such as intentionality) to be of greater importance than that of person and situation (e.g., Fletcher, 1983; Lewis, 1995; Passer, Kelley, & Michela, 1978). Likewise, developmental research, which has not relied on the person/situation assumption, found that children's emerging explanations of behaviour centre on the distinction between intentional and unintentional behaviour and on the understanding that intentional behaviours reflect the agent's goals and beliefs (e.g., Bartsch & Wellman, 1995; Wellman, Hickling, & Schult, 1997).

If there is no empirical evidence for PS, is there at least good theoretical reason to believe in PS? Unfortunately not. The person/situation assumption is not derived, for example, from a model of people's conceptual framework of behaviour. In fact, it runs counter to pertinent analyses in the philosophical literature on human action and action explanation, which distinguish between intentional and unintentional behaviour and identify a unique mode of explaining intentional behaviour in the form of the agent's reasons (e.g., Davidson, 1963; Mele, 1992; Mischel, 1969; Searle, 1983). A simple example should suffice for now to illustrate that the person/situation dichotomy simply does not capture the nature of people's explanations of intentional action. Consider this scenario:

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

Pauline now tries to explain Fred's action of leaving the note in her mailbox. Kelley's attribution model would claim that Pauline's choice is between a person attribution (something about Fred caused the action) and a situation attribution (something about herself or the circumstances caused the action). But right away this is a confusing choice. Surely something about Fred must have been causing the

action (e.g., his intention, his motivation) if his putting the note in her mailbox was intentional. And, of course, the situation figured into the action as well, or at least the situation as seen by Fred; perhaps he thought Pauline would like to have some company, or he expected her to be an ideal collaborator. What the search for person/situation attributions misses entirely is what the explainer actually does when faced with a scenario like this. Pauline will simply try to find out Fred's *reasons* for leaving the note – his goals, beliefs and assumptions. A theory of behaviour explanation must incorporate the concept of reasons into its theoretical repertoire.

Finally, what is the historical basis for PS? Here we encounter two major misunderstandings. To begin with, Lewin (1936), as one historical source of the person/situation dichotomy, meant it as a sketch of the *reality* of social behaviour – that scientists can start out with the assumption that behaviour is a function of the person and the situation, including all their complex interplays. But Lewin at no point argued that *ordinary people* see social behaviour in terms of person and situation causes.

What is perhaps more surprising is that Heider (1958), the most widely cited historical source for PS, also did not claim that lay people divide the world into person and situation causes. Instead Heider argued that, when trying to explain events in the social world, people make a fundamental distinction between "personal causality" and "impersonal causality". What he referred to in using these terms are distinct causal models that ordinary people bring to social perception (Heider, 1958, pp. 100–101). The personal causal model is applied to the domain of intentional behaviour, for which people assume the involvement of an intention as the critical force that brings about the action. The impersonal causal model is for all other domains (i.e., unintentional human behaviour as well as physical events), in which causes simply bring about effects – without any involvement of intentions.

The confusion between Heider's distinction of personal and impersonal (or intentional and unintentional) causality on the one hand and the traditional person/situation dichotomy on the other was not just a curious historical accident[2]; it had rather significant theoretical consequences. Attribution theories after Heider ignored the intentional/unintentional distinction and built models that applied alike to all behaviours. But it was precisely Heider's (1958) point that not all behaviours are explained alike. He specifically stated that, whereas unintentional behaviours were explained simply by causes, intentional actions were explained by the "reasons behind the intention" (Heider, 1958, p. 110; see also pp. 125–129). Even in 1976, around the peak of attribution research, Heider observed that explanations of intentional action by way of reasons had not been adequately treated in contemporary attribution work (Ickes, 1976, p. 14). Sadly, nothing seems to have changed in this regard, if we take social psychology textbooks and major surveys of attribution research as barometers (e.g., Anderson, Krull, & Weiner, 1996; Aronson, Wilson, & Akert, 2002; Försterling, 2001; Gilbert, 1998).

Might social psychology have held on to the simplified model of person/situation attribution because for a long time there was no alternative available? This cannot be quite right, because alternative viewpoints have been voiced repeatedly (e.g., Buss, 1978; Lalljee & Abelson, 1983; Locke & Pennington, 1982; Read, 1987; White, 1991). It is true, however, that these alternative viewpoints did not resolve the contradictions between the various models and did not provide an integrative theory of behaviour explanation. Such an integrative theory is what I hope to offer in this

chapter, but first I must briefly discuss the second core claim of traditional attribution theory.

Kelley's (1967) claim that covariation analysis underlies the construction of lay explanations – claim COV – is problematic as well. First, the covariation claim is poorly supported empirically. The available evidence shows that people can make use of covariation information when it is presented to them by the experimenter (e.g., Försterling, 1992; McArthur, 1972; Sutton & McClure, 2001; Van Kleeck, Hillger, & Brown, 1988). But there is no evidence that people spontaneously search for covariation information when trying to explain behaviour. In fact, very few studies have examined whether and when people actively seek out covariation information in natural contexts. In a rare exception, Lalljee, Lamb, Furnham and Jaspars (1984) asked their participants to write down the kind of information they would like to have in order to explain various events, and covariation information was in low demand under these conditions. A few additional studies examined people's choices between receiving covariation information and some other information, and there too explainers were less interested in covariation information than in information about generative forces or mechanisms (Ahn, Kalish, Medin, & Gelman, 1995).

Sutton & McClure, 2001 NOT IN REFS

The theoretical foundation for claim COV is dubious as well. The notion of covariation analysis was a creative analogy to scientific and statistical reasoning, but it was not grounded in any model of human inference. The covariation thesis also contradicts what we know about behaviour explanations in communicative contexts (Hilton, 1990; Kidd & Amabile, 1981; Turnbull & Slugoski, 1988). In constructing explanations for another person (the audience), the speaker's choice of a particular causal factor is guided far less by covariation analysis than by impression management (i.e., selecting a cause that puts the agent or explainer in a certain evaluative light; Tedeschi & Reiss, 1981) and by audience design (i.e., selecting a cause that satisfies the listener's curiosity or expectation; Slugoski, Lalljee, Lamb, & Ginsburg, 1993). So even if there are some contexts in which covariation analysis is important, it is clearly not the only cognitive process by which explanations are constructed (Malle, 2004).

Apart from the lack of support for its two core claims, traditional attribution theory and its successors have two additional limitations. For one thing, they treat explanations as a purely cognitive activity, so there is no accounting for such social functions of explanation as clarifying something for another person or influencing an audience's impressions. Moreover, traditional attribution theory does not specify any psychological factors (besides raw information) that influence the construction of explanations. Specifying these factors would allow us to predict such important phenomena as actor/observer asymmetries, self-serving biases and the like.

Demands on a New Theory of Explanation

The difficulties with standard attribution theory imply a number of desirable features that a new theory of explanation must have. First, instead of allowing a reduction to person and situation causes, the theory has to capture the concepts that

actually underlie people's thinking and reasoning about human behaviour, such as agency and intentionality.

Second, the new theory must identify additional cognitive processes, besides covariation analysis, that are recruited to construct explanations. It should also begin to specify the conditions under which each of these cognitive processes is used.

Third, the theory has to integrate the social-communicative aspect of explanations with the cognitive aspect. It must be made clear in what way the social and the cognitive nature of explanation are tied together and in what respects they differ.

An Alternative: The Folk-Conceptual Theory of Explanation

The theory of behaviour explanation that my colleagues and I have developed appears to meet the above demands and may be able to supersede attribution theory as an account of people's behaviour explanations (Malle 1999, 2001, 2004; Malle, Knobe, O'Laughlin, Pearce, & Nelson, 2000). I call it the *folk-conceptual* theory of behaviour explanation because its basic assumptions are grounded in people's folk concepts of mind and behaviour.

The theory has three layers. The first layer concerns the conceptual framework that underlies behaviour explanations (and helps meet the first demand specified above). The starting point is Heider's insight that people distinguish sharply between intentional and unintentional behaviour and conceptualize these two behaviours very differently (Malle, 2001; Malle & Knobe, 1997a). As I will show in more detail, this conceptualization implies three distinct modes of explaining intentional behaviour, along with a fourth mode of explaining unintentional behaviour and distinct explanation types within each mode (Malle, 1999) – for example, the mode of reason explanations breaks down into belief reasons, desire reasons and valuings.

Only Malle & Knobe, 1997 in refs

The next layer of the theory concerns the psychological processes that govern the construction of explanations (and helps meet the second and fourth demands above). In constructing explanations, people have to solve two different problems. The first is to choose among the various explanatory tools (i.e., modes and types of explanations), and three factors appear to determine those choices: features of the behaviour to be explained (e.g., intentionality, difficulty), pragmatic goals (e.g., impression management, audience design) and information resources (e.g., stored information, perceived action context). The second problem in constructing explanations is that people must select *specific* reasons, causes and so on (not just "a belief reason" or "a situation cause"), and they do so by relying on a number of cognitive processes separately or jointly (e.g., retrieving information from knowledge structures, simulation, projection, rationalization and, occasionally, covariation analysis).

The third layer of the theory is a linguistic one that identifies the specific linguistic forms speakers have available in their language to express behaviour explanations (this layer helps meet the third and fourth demands above). People can exploit

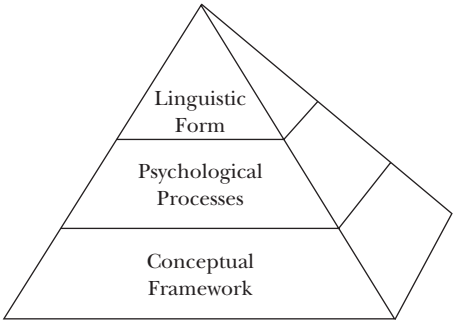

FIGURE 1.1
Three layers of the folk-conceptual theory of behaviour explanation

these linguistic forms when using explanations as a tool of social influence, such as to distance oneself from an agent's reason (e.g., "Why did she refuse dessert?" "Because she's been gaining weight" vs. Because *she thinks* she's been gaining weight"; Malle et al., 2000).

The three layers can be depicted in a hierarchy (see Figure 1.1) that considers the conceptual framework as the foundation, the psychological processes as operating on this foundation, and the linguistic layer as operating on both layers underneath. I now develop the first two layers in detail and report supporting empirical evidence.

The Conceptual Framework of Behaviour Explanations

Traditional attribution theory postulated a simple set of concepts that were supposed to underlie lay behaviour explanations. There were "effects" (behaviours, outcomes or events) and "causes," and the latter were classified into personal (dispositional or internal) causes and situational (external) causes.[3] This framework is incompatible with what we know about children's emerging theory of mind and behaviour – the conceptual network that 4- to 5-year-olds rely on when interpreting and thinking about human behaviour (Gopnik & Meltzoff, 1997; Perner, 1991; Wellman, 1990). In children's theory of mind we see the importance of a concept of intentionality, of mental states contrasted with observable behaviours and of specific mental states, such as intentions, beliefs and desires, that are used to explain intentional behaviour. It is rather unlikely that people forget these concepts and distinctions when they grow up and instead explain behaviour using a person/situation dichotomy. But even though the attribution framework was criticized repeatedly for its omission of mental-state concepts such as reasons, goals or motives (e.g., Buss, 1978; Read, 1987; White, 1991), no comprehensive revision of the attribution framework has yet been offered.

The folk-conceptual theory of explanation takes seriously the complex network of assumptions and distinctions that underlie people's thinking about behaviour – whether in early childhood or adulthood – and thus integrates important concepts such as reasons and goals into a revised model of how people explain behaviour.

Intentionality

The first conceptual postulate of the theory is that when people deal with human behaviour, they distinguish sharply between intentional and unintentional behaviour (Heider, 1958; Malle, 1999; White, 1991). Social perceivers show a high level of agreement (α = .99) in their intentionality judgements (Malle & Knobe, 1997a, Study 1), and they do so by relying on a shared folk concept of intentionality. This concept normally includes five requirements for an action to be judged as intentional: The action must be based on a desire for an outcome, beliefs about the action's relationship with this outcome, a resulting intention to perform the action, and skill and awareness when actually performing it (Malle & Knobe, 1997a, Studies 2–4).

Only Malle & Knobe, 1997 in refs

Subsequent studies showed that an intention is seen as a commitment to act that flows from a reasoning process (Malle & Knobe, 2001) in which the agent weighs a number of beliefs and desires and settles on a course of action.

Only Malle & Knobe, 1997 in refs

Reasons versus Causes

The second conceptual postulate is that people explain intentional behaviour differently from the way they explain unintentional behaviour.[4] Specifically, whereas unintentional behaviour is explained by causes, intentional behaviour is explained primarily by reasons (Buss, 1978; Davidson, 1963; Donellan, 1967; Locke & Pennington, 1982; Malle, 1999; Mele, 1992; Read, 1987; Searle, 1983). Reasons and causes are both seen as "generating factors", but reasons are a unique kind of generating factor. They are representational mental states – that is, states such as beliefs and desires that represent a specific content (what is believed or what is desired). For beliefs or desires to be the agent's reasons, their content had to be (in the explainer's eyes) part of a reasoning process that led the agent to her decision to act. When an explainer claims, "Anne invited Ben to dinner because he had fixed her car," then the explainer must presume[5] that Anne actually considered Ben's fixing her car and *for that reason* invited him to dinner. The notion of a reasoning process does not require that the agent go through an extended full-fledged deliberation, but it does require, in people's folk theory of mind, that the agent (a) considered those reasons when deciding to act and (b) regarded them as grounds for acting. These two conditions, which I have called *subjectivity assumption* and *rationality assumption* respectively (Malle, 1999, 2004), define what it is to be a reason explanation. It is not enough for some mental states to be general grounds for acting; if the agent did not actually consider them when deciding to act, they were not the agent's reasons (Malle et al., 2000). Likewise, it is not enough for some mental states simply to be on the agent's mind while she is deciding to act; if she did not regard them as grounds for her acting, she did not act for those reasons.

Causes of unintentional behaviour, of course, do not have to meet a subjectivity or rationality requirement. They can be unconscious or irrational; all that counts is that they are presumed to be factors that brought about the behaviour in question. Because unintentional behaviour presupposes neither intention nor awareness on the part of the agent, the way by which causes bring about unintentional behaviour is independent of the agent's reasoning and will. In that sense, causes are "impersonal", as Heider (1958) put it.

As an illustration of the difference between reasons and causes, consider the following two explanations:

1. Kim was nervous about the math test because she wanted to be the best in class.
2. Kim studied for the math test all day because she wanted to be the best in class.

In the first case, the desire to be the best caused Kim's nervousness, but that desire did not figure as part of a reasoning process, nor did Kim regard it as grounds for being nervous. In fact, it is possible that Kim was not even aware of her desire to be the best. The situation is very different for the second case. To understand this reason explanation is to assume that Kim decided to study in light of her desire to be the best and regarded such a desire to be grounds for studying all day.

Other Modes of Explaining Intentional Behaviour

Reasons are the default mode for explaining intentional actions, making up about 70% of these explanations (Malle, 2004). But at times people use one of two alternative explanation modes. One is to explain actions not with the agent's reasons but with factors that preceded those reasons and presumably brought them about (see Figure 1.2). Whereas reasons capture what the agent herself weighed and considered when deciding to act, causal history explanations capture the various causal factors that led up to the agent's reasons. These "causal history of reason", or CHR, explanations literally describe the causal history, origin or background of reasons (Malle, 1994, 1999; see also Hirschberg, 1978, Locke & Pennington, 1982), and such a history could lie in childhood, cultural training or traits, or in situational cues that triggered, say, a particular desire.

If we wanted to offer a CHR explanation for Kim's studying for a test all day, we might say, "She is achievement-oriented" or "She comes from a family of academics" or "That's typical in her culture." CHR explanations can cite something about the agent or something about the situation, so the "locus" of the causal history factor can vary considerably. What fundamentally defines an explanation as a causal history explanation is (a) that it explains an intentional action, (b) that it clarifies why

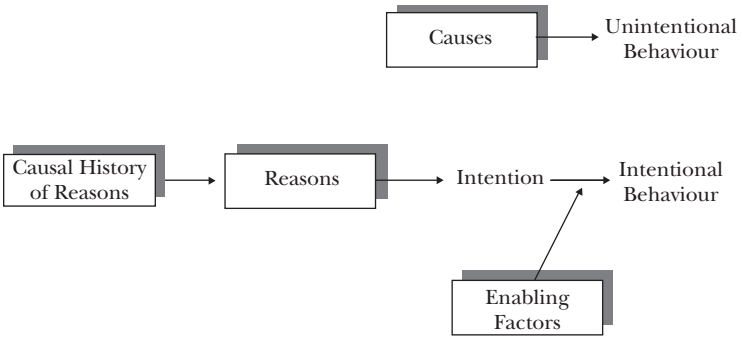

FIGURE 1.2
Four modes of folk explanation, with arrows indicating presumed causal connections

the person decided to act as described and (c) that it is not a reason explanation. Condition (c) entails that neither the subjectivity nor the rationality assumption holds for CHR explanations. In the eyes of the explainer, Kim did not reason, "I am achievement-oriented, therefore I should study all day" or "It's typical in my culture to study all day, so I will too." Causal history factors exert their causal power regardless of the agent's awareness of those powers, and even though they can trigger the agent's reasoning, they are not part of it. In many cases we can construct a chain such that the agent's reasons are what moved her to act and the CHR factors are what brought about or strengthened those reasons (see Figure 1.2). It is likely, for example, that the desire reason "to be the best" is a result of one's being achievement-oriented or coming from a family of academics.

There is a second alternative mode for explaining intentional behaviours. This one does not clarify what *motivated* the agent to act (as reasons and causal histories do) but rather what enabled the action to succeed; hence we call it an *enabling-factor* explanation. Whereas reason explanations and CHR explanations explain both the agent's action and her intention to act (even before she implements the action), enabling-factor explanations apply only when the action was actually completed, and they clarify how it was possible that the action succeeded. The explainer thus cites important causal factors (e.g., abilities, effort, opportune circumstances) that presumably helped the agent turn the intention into a successful action (see Figure 1.2). For example, if we wonder how it was possible that Kim indeed studied all day for the math test, we might say, "She had made a big pot of coffee." Likewise, if we wondered how she could complete the eventual test in just 15 minutes, we might say, "She had one of those new calculators" or "She worked very efficiently."

People's conceptual toolbox for explaining behaviour thus contains four modes of explanation: one for unintentional behaviour (causes) and three for intentional behaviour (reasons, causal histories and enabling factors). These modes of explanation can be reliably discriminated when coding naturally expressed behaviour explanations (with inter-rater reliabilities of $\kappa > .80$; e.g., Malle et al., 2000). Importantly, a small set of psychological processes determines the conditions under which each mode occurs (Malle, 2004), to which I will return shortly.

The Nature of Reasons

The fourth conceptual postulate of the theory of folk explanations consists of a set of specific claims about the types and features of reasons. As already mentioned, reasons are representational mental states. We can therefore distinguish between the specific mental state that is cited in the explanation and the content of that state, and this distinction yields three classifications of reason features.

1. On the mental-state side, three types of states can function as reasons: beliefs (e.g., knowing, thinking), desires (e.g., wanting, needing, trying) and valuings (e.g., liking/disliking, enjoying).
2. On the content side, we can classify what is believed, desired or valued into categories such as desirable versus undesirable, or into the traditional person/situation categories (e.g., "because he wanted a car" [situation]; "because he wanted to be rich" [person]), or into alternative.

[WORDS MISSING?

TABLE 1.1

Reasons in Their Marked and Unmarked Form

Behaviour	Reason Type	Marked Form	Unmarked Form
Why did they sell their car?	Belief	They felt it was too small for the family.	It was too small for the family.
Why did he go to the coffee shop?	Desire	He wanted to have a real Italian espresso.	To have a real Italian espresso.
Why did she stay until after 10:00?	Valuing*	She liked the show.	The show was fun.

* Among valuings, unmarked forms are extremely rare. Moreover, the unmarked forms cannot be created by omitting the mental-state verb (as with most beliefs and desires); instead, unmarked valuings are expressed by an evaluative claim about the content of the reason (e.g., "It's fun!").

3. Once more on the mental-state side, but at the level of linguistic form, reasons can be expressed either with mental-state markers – verbs such as "think", "believe", "want", "need" or "like" that indicate explicitly what kind of mental state the reason represents – or without such markers. Table 1.1 exemplifies reason explanations that differ in their mental-state type and are in either marked or unmarked form.

When we examine reasons in detail, then, we find three features by which reasons can differ: the type of mental state the reason reflects, the content of that reason, and the presence or absence of mental-state markers. But what psychological functions and properties are associated with each of these three features?

The choice between belief reasons and desire reasons has at least two psychological properties. First, belief reasons are favoured over desire reasons when the explainer tries to present the agent in a rational light (Malle et al., 2000). Beliefs seem to highlight the agent's rational deliberation and (if left unmarked) the "objective facts" on which that deliberation was based. Second, belief reasons (compared to desire reasons) are more often used by actors than by observers. Our data so far suggest that this asymmetry may be the result of actor/observer differences in both impression management and information resources (Malle, Knobe, & Nelson, 2004). Specifically, actors are typically more engaged than observers are in presenting the action in a rational, positive light (for which belief reasons are particularly suitable), and actors sometimes know more than observers do about the specific considerations that led to the specific action (which are often more suitably expressed by belief reasons).

Mental-state markers may appear to be a minor linguistic variation, but they have several important properties as well. Pairs of marked and unmarked explanations, such as those in Table 1.1, not only differentiate the two ingredients of reasons (the type of mental state the reason represents and the content of that reason), they also reveal that reason explanations conceptually refer to the agent's mental states even when their linguistic surface does not explicitly mention such a state. That is, an explanation such as "My father never lets us go out *because something might happen to us* refers to the father's belief that something might happen to his daughters, even though that belief isn't mentioned. Without the theoretical categories we have proposed, one might falsely assume that the explanation refers to an objective situation cause ("something happening") that could be subsumed under the traditional

person/situation categories. But nothing has actually happened at the time of the explanation, so the objective situation itself cannot be the cause of the father's action. Instead, the father *believes* that something might happen, and everybody who hears or reads the explanation will infer that it is this belief that explains the father's action.

Why might explainers omit mental-state markers? For one thing, omitting a mental-state marker makes the reason sound more objective and true. By stating that "something might happen," the daughter refers to a potential reality that serves to justify the father's action. Conversely, adding a mental-state marker allows explainers to distance themselves from an agent's reason. By stating that the father never lets them go out "because he *thinks* something might happen," a sceptical observer would indicate disagreement with the father's belief and cast some doubt on its plausibility.

Consider another example of distancing behaviour, which we presented to 91 undergraduate students (Malle et al., 2000, Study 6). Cliff and Jerry are at a dinner party. Cliff asks Jerry, "Why did your girlfriend refuse dessert?" Jerry responds by saying either "She thinks she's been gaining weight" (marked belief) or "She's been gaining weight" (unmarked belief). After reading the vignette, participants rated (on a scale from 0 to 8) how happy Jerry was with his girlfriend's current weight. As predicted, 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 ($M = 2.6$), $F(1, 88) = 21.9$, $p < .01$, $\eta^2 = 20\%$.[6]

Reason contents, finally, have not proven to carry any clear psychological function, at least as long as they are classified according to the traditional person/situation categories. For example, actors and observers do not differ in their content of reasons (Malle et al., 2004), explanations of group actions do not differ from explanations of individual actions in the content of reasons (O'Laughlin & Malle, 2002), and impression management does not have a reliable impact on reason contents. There may well be a psychological function associated with other aspects of reason contents (such as their social desirability), but this possibility remains to be investigated.

In sum, reasons have a complex conceptual and linguistic structure that is not reducible to any traditional causal categories. To understand reasons is to understand their nature as representational mental states and their resulting three features: the type of mental state they reflect (belief, desire, valuing), the content of that state, and the linguistic form as being marked or unmarked. Current evidence suggests that at least two of these features are associated with important psychological functions or processes.

Types of Causal Factors

The fifth conceptual postulate of the theory of folk explanation concerns the types of causes, causal histories and enabling factors that people construct. These explanatory modes all refer to causal factors that can in principle be classified by their locus, following the person/situation dichotomy in traditional attribution research. However, that dichotomy has been ambiguous in that the person category sometimes referred to stable traits, usually labelled "dispositions", whereas at other times it referred to all causal factors internal to the agent, whether stable or not. A more precise way to classify these causal factors is to use the label *person* as an overarching

category that refers to all causal forces inside the agent and to reserve the word *trait* for person factors that are stable parts of the agent's personality. That way we break up the causal forces into two orthogonal contrasts: person versus situation and, among person factors, traits versus non-traits.[7] Using this finer classification we have found that actors and observers differ in their use of trait versus non-trait person causes, provided that the observer knows the agent well. By contrast, actors and observers do not differ in their use of person-versus-situation causes, nor does any predictive force emerge from either the person/situation or the trait/non-trait classification in studies comparing explanations of groups and individuals (O'Laughlin & Malle, 2002) or impression management (MacCionnaith, 2003; Malle et al., 2000).

These postulates about the folk-conceptual structure of behaviour explanations paint a rich picture of behaviour explanations in which explainers have to choose (consciously or not) between multiple modes of explanations, different types within each mode, and alternative linguistic forms. Formulations of traditional attribution theory that succeeded Heider (1958) failed to distinguish between all these different explanatory tools, confounding intentional and unintentional behaviour, collapsing four distinct modes of explanation into one unitary causal attribution and ignoring many finer-grained types and forms of explanation. Moreover, whereas many of these modes, types and forms of folk explanations have predictable psychological significance (e.g., in actor/observer asymmetries, group/individual differences and impression management), the traditional attribution categories of person/disposition and situation show very limited predictive power, because they do not carve up, as it were, explanatory phenomena at their joints.

The question I now turn to constitutes the second layer of the folk-conceptual theory of explanation: which psychological processes guide the choice among the multifaceted explanatory tools.

Psychological Processes

When examining the psychological processes that guide folk explanations of behaviour, we have to separate two problems that folk explainers face. For one thing, they must choose a tool from the large toolbox of explanatory modes, types and forms (e.g., a marked belief reason, a trait-enabling factor). In addition, they must provide a *specific* instance of any explanatory tool. No ordinary explanation stops at the level of conceptual categories – one cannot explain an everyday behaviour by saying "She had a reason" or "There was some trait-enabling factor." Instead, folk explanations of behaviour must be tailored to the agent, action and context so that an action such as "She moved all the furniture" is explained by a specific reason such as "because she expects a lot of people for the party" or by a specific trait such as anxiousness.

Of these two problems – the choice of an explanatory tool and the construction of a specific explanation – the first is scientifically more tractable. In fact, it is unlikely that a psychological theory will ever predict the precise contents that explainers provide in context-specific explanations. But the cognitive process of *searching for* such contents may well be predictable, and this issue promises to be an

intriguing domain for future research. Before I sketch this future research, however, I summarize what we know about the first problem, the choice of explanatory tools.

Determinants of Choosing Explanatory Tools

Our research findings on people's differential use of explanation modes, types and forms are best accounted for by three factors: attributes of the explained behaviour, pragmatic goals and available information resources.

Behaviour Attributes

Before explaining a given behaviour, social perceivers make several (often implicit) judgements about that behaviour. To begin, they consider the behaviour's *intentionality* and, as a result, select distinct modes of explanation. If judged unintentional, a behaviour is explained by causes; if judged intentional, it is explained by reasons, causal histories or enabling factors. To illustrate, in one study a group of participants made intentionality judgements for twenty behaviours, whereas a second group of participants offered explanations for those same behaviours. Analyses showed that the greater the judged intentionality of a given behaviour, the greater the probability that people gave reason explanations ($r = .91$) and the smaller the probability that they gave cause explanations ($r = -.90$; Malle, 1999, Table 1).

A second important behaviour attribute is the perceived difficulty of intentional actions. If an action is considered difficult to accomplish, explainers will often provide enabling factors. If it is not difficult, they are apt to choose reasons or causal history explanations (Malle et al., 2000; McClure & Hilton, 1997, 1998).

A third attribute is whether the social perceiver explains a singular behaviour or a behaviour trend (across time or agents). If the behaviour is judged to be a trend, the rate of CHR explanations increases significantly compared to singular behaviours (O'Laughlin & Malle, 2002). That is because each behaviour within the trend may have a different reason explanation, and citing all of those reasons would be extremely cumbersome. One or two causal history factors may suffice to indicate the background that triggered the full array of differential reasons. For example, a mother who was asked to explain why she went shopping many times a week answered this way: "Because I have three children." The trend of actions in question is parsimoniously explained by offering the causal history of having three children, because it underlies the variety of specific reasons she has for shopping each time (e.g., buying more milk, a new supply of diapers or a carpet cleaner for crayon stains).

Pragmatic Goals

When social perceivers explain behaviours in communicative contexts, they have a variety of smaller or larger goals they try to accomplish with their explanations, such as to lessen another person's confusion, manage their own status in the interaction, or fend off blame. Two kinds of goals can be distinguished. In *audience design*, the explainer tailors the explanation to the audience's needs and existing knowledge (Hilton, 1990; Slugoski et al., 1993). A clear case of audience design is when the explainer matches an explanation mode to the type of question asked (Malle et al.,

2000; McClure & Hilton, 1998). Specifically, the question by which someone requests an explanation can either inquire about the agent's immediate motivation ("What did she do that for?") and thereby demand a *reason* explanation; it can inquire about the background of that motivation ("How come?") and invite a *causal history* explanation; or it can inquire about the factors that enabled a successful action outcome ("How was it possible that she did that?"), demanding an *enabling-factor* explanation.

The second pragmatic goal, *impression management*, engages the explainer in an act of social influence, using the behaviour explanation to create certain beliefs, perceptions or actions in the communication partner. For example, people increase their use of causal history explanations when accounting for negative actions (Nelson, 2003), they increase their use of belief reasons when trying to appear rational (Malle et al., 2000), and they explicitly add a mental-state marker to their belief reasons when they want to distance themselves from the agent (e.g., "Why is he looking at apartments?" "He thinks I am moving in with him"; Malle et al., 2000).

Information Resources

Different explanation modes and types have different information demands. Reason explanations, for example, require relatively specific information about the agent, the behaviour and the context, whereas causal history explanations and enabling-factor explanations may get by with less context-specific information. Similarly, belief reasons often require fairly idiosyncratic information about the agent's deliberations, whereas desire reasons can sometimes be constructed from the nature of the behaviour alone. In support of this difference between desire and belief reasons, we found that observers normally provide fewer belief reasons (and more desire reasons) than actors do. However, when observers know the agent well, their rate of belief reasons increases to equal that of actors (Malle et al., 2004).

The three determinants of explanatory choice, along with the conceptual nature of modes and types of explanations, jointly provide the theoretical basis to predict and account for a number of important phenomena. So far we have successfully applied this approach to predicting strategies of impression management in explanation (Malle et al., 2000), predicting differences between explaining group and individual behaviours (O'Laughlin & Malle, 2002), and predicting a variety of actor/observer asymmetries in explanations of behaviour (Malle et al., 2004). Other domains are open to investigation as well, such as close relationships, negotiation, psychopathology and cross-cultural explorations (see Malle, 2004).

Constructing Specific Explanations

The process of constructing explanations with a specific content has remained largely unexplored in 40 years of attribution research. One reason for this omission was that standard attribution models tried to predict only whether an explainer would give a "person attribution" or a "situation attribution." When one describes explanatory work at such a general level, the process of constructing specific explanations simply does not come up (cf. Kruglanski, 1979). Another reason for this

omission was the assumption that explainers use only one cognitive process to arrive at their attributions, namely, covariation analysis. Unfortunately, this assumption was never adequately supported, as all such tests showed merely that people could *respond to* covariation information if it was presented by the experimenter. The few studies that examined whether people spontaneously search for covariation information in more natural contexts cast serious doubt on the ubiquity of covariation analysis (Ahn et al., 1995; Lalljee et al., 1984).

Several critics of standard attribution theory have proposed cognitive processes that exist alongside covariation analysis and help the explainer construct specific explanations. First, explainers recruit event-specific, agent-specific or general *knowledge structures* (Abelson & Lalljee, 1988; Ames, 2004; Lalljee & Abelson, 1983; Read, 1987). Second, they use the two related processes of *simulation* (imaginative representation of the agent's mental states; Goldman, 1989, 2001; Gordon, 1986, 1992; Harris, 1992) and *projection* (assuming that the agent's mental state is the same as one's own; Ames, 2004; Krueger & Clement, 1997; Van Boven & Loewenstein, 2003). In addition, two principles direct the explainer's knowledge recruitment and simulation: (a) the "method of difference", which contrasts the event in question with an alternative event and tries to identify the critical difference (e.g., Cheng & Novick, 1992; Hilton & Slugoski, 1986; Kahneman & Miller, 1986; McGill, 1989), and (b) a premium on identifying generative forces or mechanisms (Ahn et al., 1995; Ahn & Kalish, 2000; Cheng, 2000; Johnson, Long, & Robinson, 2001).

Elsewhere I applied this set of proposed processes to naturally occurring explanations and developed hypotheses about the relationship between particular processes and particular explanatory tools (Malle, 2004). Because of space constraints I can only summarize the main results of this exploration, grouped by the four modes of folk-behaviour explanations.

Cause Explanations

From the observer perspective, when the explainer has little familiarity with the agent and/or did not directly observe the unintentional behaviour, reliance on general knowledge, including stereotypes and cultural scripts, is likely to be high (Ames, 2004). As familiarity increases, and especially when the explainer directly observes the behaviour, the use of simulation and projection will increase. From the actor perspective, stored knowledge will be dominant (e.g., recall of events immediately preceding the unintentional behaviour), but for private wonderings about recurring and puzzling experiences, covariation analysis may be recruited (e.g., when one wonders about a recurrent headache).

Causal History Explanations

For both actor and observer perspectives, the predominant process in generating CHR explanations is the recruitment of knowledge structures relevant to the context, the agent or the action. Simulation processes may come into play when an observer searches for specific causal history factors in the agent's experiences, and covariation analysis will become dominant when the explainer (from either perspective) searches for a common causal history behind a trend of actions.

Enabling-Factor Explanations

The construction of enabling-factor explanations also relies primarily on specific or generic stored knowledge (e.g., about the kinds of facilitating forces that enable particular actions), but in achievement domains (e.g., grades, sports victories), covariation analysis can become relevant as well. Simulation is largely absent when constructing enabling-factor explanations because people cannot easily simulate abilities, opportunities or other facilitating forces.

Reason Explanations

When selecting reason explanations, actors never use covariation calculation. Instead, they typically have (or believe they have) access to the reasons that initially prompted their intention, relying on a process of direct recall (Brewer, 1994). This process becomes less important, and general knowledge structures more important, when the action explained was fairly automatic (because the memory trace for the action's reasons is weak) or when the action was performed long ago (because the memory traces of reasons may have washed out). Also, when actors alter their explanation for impression management purposes, they will not use covariation analysis but rather will recruit those reasons from knowledge structures that would best meet their impression goals. Observers may occasionally use covariation analysis when they wonder about an agent's repeated choice among well-defined options, but in most cases of singular actions, knowledge structures and simulation will prevail.

Besides testing these hypotheses about processes involved in constructing specific explanations, future research will also have to examine the interrelationship between (a) these construction processes and (b) the earlier-mentioned general determinants of explanatory choice (behaviour attributes, pragmatic goals and information resources). In some cases the general determinants will first favour a particular mode of explanation and then elicit a construction process suitable for this mode. For example, audience design goals may directly favour causal history explanations, which are bound to be searched for in knowledge structures. In other cases, the determinants may directly favour a construction process, which in turn provides a specific explanation content. For example, limited information resources may encourage the explainer to use simulation or projection, which are likely to deliver causal histories or desire reasons.

Summary and Conclusions

For a long time the decisive word on lay behaviour explanations came from attribution theory. In this chapter I have tried to convince the reader that this decisive word was too often wrong. When dealing with explanations of unintentional behaviours and outcomes, attribution theory provides a fine conceptual framework, though some improvements were recommended here as well (e.g., the distinction between traits and other person causes, and the multiple cognitive processes that are used to construct specific explanations). Where attribution theory fails entirely is in dealing

with explanations of intentional behaviours, both at the conceptual level and at the level of psychological processes.

The folk-conceptual theory of behaviour explanation identifies the conceptual framework that underlies lay explanations of intentional behaviour, and thus focuses on the key role of intentionality and the resulting distinctions between modes of explanation (reasons, causal histories, enabling factors) and their specific features (e.g., beliefs, desires, mental-state markers). This first, conceptual layer of the new theory – directly tested and supported in recent work (Malle, 1999; Malle et al., 2000) – precisely describes the tools people use to explain behaviour and brings order to the complexity of naturally occurring explanations. This layer also unites two aspects of explanation that are often separated in the literature: explanations as cognitive (private) events and explanations as social (public) acts (Malle, 2004; Malle & Knobe, 1997b). Despite their different implementations, antecedents and consequences, these two kinds of explanations are built from the same conceptual framework that specifies the modes and types of explanation that are available to both private and public explainers.

Only Malle & Knobe, 1997 in refs

A second layer of the folk-conceptual theory concerns the psychological processes that give shape to explanations as cognitive and social acts. For one thing, three primary psychological determinants (judged behaviour features, pragmatic goals and information resources) guide people's choices of modes and features of explanation. Moreover, explainers must select specific contents of explanations in specific situations, and they do so by relying on a variety of cognitive processes, including knowledge structures, direct recall, simulation and covariation analysis.

The folk-conceptual theory not only describes but also accounts for many of the regularities of behaviour explanations, including the conditions under which various explanatory tools are used, as well as the social functions they serve (e.g., Malle, 1999; Malle et al., 2000). As a result, we have been able to show that the concepts and distinctions in this new theory have predictive power when it comes to investigating impression management (Malle et al., 2000), asymmetries between individual and group targets (O'Laughlin & Malle, 2002), asymmetries between actor and observer perspectives (Malle et al., 2004) and self-servingness (Nelson & Malle, 2004). In contrast, whenever we analysed the same explanation data in terms of a person/situation attribution model, virtually no such predictive power was found.

Previous attempts at replacing attribution theory were not very successful, and so the textbooks still focus on Kelley's covariation model and person/situation attributions by perspective, self-servingness and the like. Perhaps this is because the person/situation distinction is so seductively simple and so easily researched in the lab (using a pair of rating scales) that the field has been reluctant to abandon it in favour of any alternative, especially a more complex model of explanation. But good science must study the phenomena as they exist, and failing to do so is where the attribution approach's greatest weakness lies. In imposing a conceptual framework on people's folk explanations that is simply not people's own framework, much of attribution research has provided data that are simplified, that are difficult to interpret and that have led to false conclusions.

But how can it be that previous research on attribution phenomena failed to uncover its own limitations? This failure may have arisen from two methodological biases. First, participants were typically asked to express their explanations on predefined person/situation rating scales rather than in the form of natural verbal

utterances. As a result, people had to transform their complex explanatory hypotheses into simple ratings, which probably invited guessing strategies as to how the ratings were to be interpreted and led to severe ambiguities in the ensuing data. A high "person" rating, for example, could have indicated a confident judgement of intentionality, a reason explanation, a person causal history factor, and much more.

Second, in the few cases in which free-response explanations were analysed, the coding was greatly limited by the presupposed person/situation categories, which picked up no more than trends in the linguistic surface of explanations, such as in the use of mental-state markers (McGill, 1989; Nisbett et al., 1973; for evidence and discussion, see Malle, 1999, Study 4, and Malle et al., 2000, Study 4).

These serious problems of ambiguity and misinterpretation apply primarily to attribution research that examined intentional behaviour. In failing to distinguish intentional from unintentional behaviour and subsuming explanations for both behaviour types under the person/situation dichotomy, the traditional analysis of intentional behaviour explanations was profoundly distorted. Such distortion can be illustrated with the following example (cf. Antaki, 1994): "Why are you going to Iceland for your holidays?" "Because it's cool." The standard attribution treatment of such an explanation would be to call it a situation cause. However, the cool weather in Iceland would hardly cause the agent from afar to go there. Rather, the agent *thinks* that it is pleasantly cool in Iceland, and that is her (belief) reason for going there. (It would still be her belief reason even if it were in fact warm in Iceland.) Instead of trying to diagnose the locus of some "cause" in the vague space between person and situation, we have to recognize that folk explanations of intentional behaviour typically refer to the agent's mental state in light of which and on the grounds of which they acted.

NOT IN REFS

Past research that was focused entirely on unintentional events remains largely valid. For example, the analysis of self-serving biases in explaining achievement outcomes does not involve reason explanations, so the person/situation dichotomy captures these explanations reasonably well. However, nothing that was found about these biases can be straightforwardly extended to explanations of intentional actions. In fact, recent research using the folk-conceptual theory suggests that when people explain intentional behaviour, the degree of self-servingness and the tools of achieving it differ significantly from the traditional picture (Nelson & Malle, 2004).

Other classic attribution findings were assumed to apply to both unintentional and intentional behaviour, such as the actor/observer asymmetry. But when, in a series of studies, we examined actor and observer explanations for both unintentional and intentional behaviours (Malle et al., 2004), very little remained valid about the traditional Jones and Nisbett (1972) thesis. The person/situation dichotomy itself shows no actor/observer asymmetries (Malle, 2005); the most consistent asymmetries held for the choice between reasons and causal histories, beliefs and desires, and marked versus unmarked beliefs; and the only finding that supported traditional claims was that observers used more trait explanations than actors did, but only for unintentional behaviour and when they knew the agent well (Knobe & Malle, 2002; Malle et al., 2004).

The shortcomings of traditional attribution theory extend to the psychological process level as well – specifically when we examine determinants of explanatory choices and mechanisms in constructing specific explanations. Many of the traditional findings in attribution research were not accounted for by reference to such

identifiable determinants as behaviour attributes, pragmatic goals or information resources. Also, the central proposition that explanations are constructed from covariation assessment (Kelley, 1967) has garnered little supportive evidence except for demonstrations that people can respond to covariation information if it is provided by the experimenter. In reality, people seem to rely on multiple psychological processes to construct explanations, including retrieval of general and specific knowledge, mental simulation and occasional covariation analysis. Exactly which processes people use for which explanation modes is an issue that has yet to be settled by empirical research, but the textbook tenet that explanations are constructed from covariation assessment is certainly inaccurate in its general form.

Despite its shortcomings, traditional attribution research has of course contributed a great deal to social psychology. It posed questions and pointed to phenomena that had simply not been considered before – among them the power of behaviour explanations (Heider, 1958; Jones et al., 1972; Quattrone, 1985); the many interesting factors that create systematic variations in explanation, such as actor/observer differences (Jones & Nisbett, 1972), self-servingness (Bradley, 1978; Heider, 1958; Miller & Ross, 1975) and impression management tactics (Tedeschi & Reiss, 1981); and the larger network of cognitive and social antecedents and consequences of behaviour explanations (Anderson et al., 1996). These impressive results and insights, however, emerged *in the context* of attribution theory, not as *predicted results* of that theory. Nothing in class attribution theory predicts that there must be actor/observer asymmetries, much less that these asymmetries be of a particular kind (Knobe & Malle, 2002). Similarly, nothing predicts explanatory tactics in impression management, a self-serving bias or other interesting phenomena. The most celebrated insights and findings of attribution research were developed in the course of attribution's research history but were never derived from a systematic theory. It is time that theoretical advances both accounted for past insights and predicted new phenomena. The folk-conceptual theory of explanation is one attempt to foster such advances.

Bradley, 1978 NOT IN REFS; Miller & Ross, 1975 NOT IN REFS

Notes

1. Because of this focus, I will not discuss Jones and Davis's (1965) correspondent inference theory. This theory has had a major impact on social psychology (Gilbert, 1998) but it does not represent a theory of behaviour explanation. For detailed arguments why it does not, see Malle (2004, chapter 1) and Hamilton (1998).
2. Part of the blame for this accident may go to Heider himself. Because his 1958 book was conceived and written over several decades, Heider was not entirely consistent in his use of terms. In one section, for example, he speaks of causes in the person and in the environment, and it seems that he actually made the PS claim (Heider, 1958, pp. 82–84). But, in fact, Heider's analysis there concerned only one particular mode of explanation – when a social perceiver tries to make sense of an "action outcome" and wonders how it could be accomplished (e.g., a weak man rowing across the river; a rookie pitcher getting twelve strikeouts in a row). In such instances the perceiver is not interested in clarifying the agent's motivation for acting but rather wonders how it was possible that the agent accomplished the desired action outcome. When explaining such accomplishments, Heider argued, the social perceiver considers two elements:

the agent's attempt to perform the action (trying) and supporting factors (can), of which some lie in the agent (e.g., ability, confidence) and some in the environment (e.g., opportunity, luck, favourable conditions). Heider thus catalogued the "conditions of successful action" (p. 110) that serve as explanations of accomplishments and are answers to "how possible?" questions. In this catalogue Heider made use of the person/situation (or internal/external) dichotomy, but he never claimed that all lay explanations of behaviour are organized around a split between person and situation causes. For further details on this historical analysis, see Malle (2004, chapter 1) and Malle and Ickes (2000).

3. Additional cause types were postulated in the domains of achievement, responsibility and depression, namely stable/unstable, specific/global and controllable/uncontrollable (Abramson et al., 1978; Weiner, 1986; Weiner et al., 1972). However, these distinctions were never clearly integrated into a theory of behaviour explanations.
4. The more precise way of speaking would be to use the term *event* instead of *behaviour*. Intentional events include not only observable actions, such as writing a letter, making a phone call or turning the radio on, but also unobservable mental states, such as deciding on a dessert, calculating a price or imagining a new carpet. Likewise, unintentional events include observable behaviours such as fidgeting, tripping or spontaneously frowning, and also unobservable mental states such as feeling sad, hearing a dog bark or having a flashback (see Malle & Knobe, 1997b; Malle & Pearce, 2001). For simplicity I use the term *behaviour* to refer to any of these events.

Only Malle & Knobe, 1997 in refs

5. In communicative settings explainers may of course lie and merely try to convince their audience that the agent acted for the reason stated. But such lying presupposes that the audience will come to believe that the agent acted for the reasons stated.
6. I should note that these effects of distancing oneself from (or embracing) an agent's reasons operate reliably only with belief reasons. Unmarked belief reasons have no surface indicator that they are a reason (and therefore can give the impression of a reality that underlies the agent's action), whereas unmarked desire reasons (e.g., "to lose weight", "so she'll stick to her diet") still have the grammatical structure of desires. As a result, unmarked desire reasons cannot easily create an impression that is different from marked desire reasons.
7. In our classifications of free-response behaviour explanations we routinely distinguish a variety of subclasses (e.g., among non-trait person factors: behaviours, mental states, group memberships, etc.) to explore any predictive validity of such classes (see Malle, 1998).

References

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

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology, 87*, 49–74.

Ahn, W., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In F. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 199–226). Cambridge, MA: MIT Press.

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

Ames, D. R. (2004). Inside the mind-reader's toolkit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology, 87*, 340–353.

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.

Aronson, E., Wilson, T. D., & Akert, R. M. (2002). *Social Psychology* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

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

Brewer, W. F. (1994). Autobiographical memory and survey research. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 11–20). New York: Springer.

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

Cheng, P. W. (2000). Causality in the mind: Estimating contextual and conjunctive power. In F. C. Keil and R. A. Wilson (Eds.), *Explanation and cognition* (pp. 227–253). Cambridge, MA: MIT Press.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99*, 365–382.

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

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

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.

Försterling, F. (1989). Models of covariation and attribution: How do they relate to the analogy of analysis of variance? *Journal of Personality and Social Psychology, 57*, 615–625.

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

Försterling, F. (2001). *Attribution: An introduction to theories, research, and applications*. Philadelphia: Psychology Press.

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

Goldman, A. I. (1989). Interpretation psychologized. *Mind and Language*, 4, 161–185.

Goldman, A. I. (2001). Desire, intention, and the simulation theory. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 207–225). Cambridge, MA: MIT Press.

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

Gordon, R. M. (1986). Folk psychology as simulation. *Mind and Language, 1*, 158–171.

Gordon, R. M. (1992). The simulation theory: Objections and misconceptions. *Mind and Language, 7*, 11–34.

Hamilton, D. L. (1998). Dispositional and attributional inferences in person perception. In J. M. Darley & J. Cooper (Eds.), *Attribution and social interaction: The legacy of Edward E. Jones* (pp. 99–114). Washington, DC: American Psychological Association.

Harris, P. (1992). From simulation to folk psychology: The case for development. *Mind and Language, 7*, 120–144.

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

Hewstone, M., & Jaspars, J. (1987). Covariation and causal attribution: A logical model of the intuitive analysis of variance. *Journal of Personality and Social Psychology, 53*, 663–672.

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

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review, 93*, 75–88.

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

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

Ickes, W. (1976). A conversation with Fritz Heider. In J. H. Harvey, W. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 3–18). Hillsdale, NJ: Erlbaum.

Johnson, J. T., Long, D. L., & Robinson, M. D. (2001). Is a cause conceptualized as a generative force? Evidence from a recognition memory paradigm. *Journal of Experimental Social Psychology, 37*, 398–412.

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

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.

Jones, E. E., Kanouse, D., Kelley, H. H., Nisbett, R. E., Valins, S., & Weiner, B. (Eds.). (1972). *Attribution: Perceiving the causes of behavior.* Morristown, NJ: General Learning Press.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 129–238). Lincoln: University of Nebraska Press.

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

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

Krueger, J., & Clement, R. W. (1997). Estimates of social consensus by majorities and minorities: The case for social projection. *Personality and Social Psychology Review, 1*, 299–313.

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

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

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

Lewin, K. (1936). *Principles of topological psychology* (F. Heider & G. M. Heider, Trans.). New York: McGraw-Hill.

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

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

MacCionnaith, K. (2003). *Accounting for actor–observer asymmetries in explanation: The role of impression management.* Unpublished senior honours thesis, University of Oregon.

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

Malle, B. F. (1998). *F.Ex: A coding scheme for people's folk explanations of behaviour.* Retrieved April 4, 2004, from darkwing.uoregon.edu/~bfmalle/fex.html

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

Malle, B. F. (2001). Folk explanations of intentional action. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 265–286). Cambridge, MA: MIT Press.

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

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

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

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33*, 101–121.

MALLE & KNOBE, 1997a and 1997b ARE CITED IN TEXT

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

Malle, B. F., Knobe, J., & Nelson, S. (2004). *Actor–observer asymmetries in folk explanations of behavior: New answers to an old question*. Manuscript under revision.

Malle, B. F., Knobe, J., O'Laughlin, M., Pearce, G. E., & Nelson, S. E. (2000). Conceptual structure and social functions of behavior explanations: Beyond person–situation attributions. *Journal of Personality and Social Psychology, 79*, 309–326.

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

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

McClure, J., & 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.

Mischel, T. (1969). *Human action: Conceptual and empirical issues*. New York: Academic Press.

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

Nelson, S., & Malle, B. F. (2004). *Self-serving biases in explanations of behavior*. Manuscript in preparation.

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

Passer, M. W., Kelley, H. H., & Michela, J. L. (1978). Multidimensional scaling of the causes for negative interpersonal behavior. *Journal of Personality and Social Psychology, 36*, 951–962.

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

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

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

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513–523.

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

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

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

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

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

Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin, 29*, 1159–1168.

Van Kleeck, M. H., Hillger, L. A., & Brown, R. (1988). Pitting verbal schemas against information variables in attribution. *Social Cognition, 6*, 89–106.

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

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

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1972). Perceiving the causes of success and failure. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 95–120). Morristown, NJ: General Learning Press.

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

Wellman, H. M., Hickling, A. K, & Schult, C. A. (1997). Young children's psychological, physical, and biological explanations. In H. W. Wellman & K. Inagaki (Eds.), *The emergence of core domains of thought: Children's reasoning about physical, psychological, and biological phenomena* (pp. 7–25). San Francisco, CA: Jossey-Bass.

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