This research was supported by an NSF CAREER award (SBR-9703315) to the first author. We thank Lou Moses for his helpful comments on a previous draft of this paper and Lew Goldberg, Robert Mauro, John Orbell, and Mick Rothbart for their suggestions on study design and analyses. Portions of this article were presented at the APA convention in San Francisco, August 1998. Correspondence should be sent to Bertram F. Malle, Department of Psychology, 1227 University of Oregon, Eugene, OR 97403, or to bfmalle@darkwing.uoregon.edu.

People's Praise and Blame for Intentions and Actions: Implications of the Folk Concept of Intentionality

Bertram F. Malle and Ruth E. Bennett
University of Oregon

According to people's folk theory of behavior, intentions are mental states directed towards future action whereas intentionality is a specific manner in which a behavior is performed, namely, with an intention and with skill and awareness. The present studies explore the implications of this distinction for how people evaluate (i.e., praise or blame) social behaviors. Two studies investigated how people's evaluations differ for intentions in actions, bare (yet unfulfilled) intentions, and intentionally performed actions. The studies also show asymmetries between blame and praise (for both intentions and actions), adding to documented negativity biases in person perception.

To make sense of social interactions, people attend to, explain, and evaluate behavior. They do so with a set of tools that has been called their "theory of mind" (Gopnik & Meltzoff, 1996; Perner, 1991; Wellman, 1990; see Moses & Chandler, 1992) or their "folk theory of behavior" (Malle, 1997; Malle & Knobe, 1997a). These tools constitute a conceptual framework within which people think about human behavior—how it is produced and how it is related to the mind. In this framework we find the concepts of intentionality and observability (Malle & Knobe, 1997b), distinct modes of explaining behavior (e.g., with reasons vs. causes; Buss, 1978; Malle, 1998), and various mental states (e.g., beliefs, desires, intentions; see Astington, 1991; Brand, 1984; D'Andrade, 1987). In this paper we focus on two such tools—the concepts of intention and of intentionality—and explore their implications for people's evaluations, in the form of praise and blame[1], of other people's behavior.

The literature has long recognized the strong impact of intentionality on praise and blame. Ever since Aristotle (1892/330 BC), scholars have described what we shall call the "intensification effect" of intentionality: If people perceive an action as intentional, their praise or blame for it is intensified compared to when they see that same action as accidental (Heider, 1958; Piaget, 1932; Schlenker, Britt, Pennington, Murphy, & Doherty, 1994; Shaver, 1985;

[1] The term *blame* has at least two meanings. The first refers to the assigning of (causal) responsibility for an outcome to a person (e.g., "People are too quick to blame their problems on others"). The second refers to the criticizing of a given intention or action (e.g., "I love you too much to blame you for going away"). We are concerned only with the second meaning, which takes *to blame* as synonymous with *to reproach* or *to criticize*.

Weiner, 1995). In recent studies, for example, Hogue and Pebbles (1997) and Kleinke, Wallis, and Stadler (1992) demonstrated that blame (as well as punishment recommendations) for rapists were intensified when the alleged rapist admitted intent.

Even though prior research has recognized the importance of intentionality for judgments of praise and blame, it has overlooked the difference between the construct of *intention* (a mental state directed towards future action) and that of *intentionality* (the manner in which the action is performed, namely, intentionally). In Hogue and Pebbles' (1997) study, for example, the experimentally manipulated portrayals of the rapist confounded (a) the presence of a prior intention (his initial intending to have sex) with (b) intentional performance (his continued advance while being aware that the woman did not want to have sex). Thus, it is impossible to tell whether the intensification of blame and punishment was due to manipulated prior intention or manipulated intentionality of performance. Indeed, the authors labeled the manipulated variable at times "intent," at times "intentionality." Similarly, Shaver (1985), in his analysis of blame and responsibility, equated a definition of intention with a definition of intentional action (p. 121), and so did Shultz (1980; 1988) in his work on children's theory of mind. Only recently have authors begun to distinguish intentions from intentionality (Astington, 1991; Meltzoff, 1995; Moses, 1993), often building on philosophical analyses (e.g., Anscombe, 1957; Bratman, 1987; Searle, 1983). Below we clarify the folk distinction between intention and intentionality and explore its relevance for the study of people's praise and blame.

Intention and Intentionality

One reason why the conceptual distinction between intention and intentionality has often been overlooked is that psychology has been lacking a systematic and validated model of people's folk concept of intentionality. Early attribution work contained merely researchers' speculations about this folk concept (Heider, 1958; Jones & Davis, 1965; see also Maselli & Altrocchi, 1969; Ossorio & Davis, 1968), and recent discussions have relied on speculative models as well (Fiske, 1989; Fleming & Darley, 1989; Shaver, 1985). Moreover, these models have disagreed in what they identify as the necessary conditions for intentional action. Malle and Knobe (1997a) sought to resolve these disagreements by examining directly how people define and use the concept of intentionality. Their empirically derived model of this folk concept shows that social perceivers distinguish between an agent's intention and the agent's performing an action intentionally. This distinction is based on at least three distinguishing features.

First, intentions and intentionality differ in the object to which they are ascribed. People treat intentions as mental states and therefore ascribe them to an *agent.* Just as the claim that a poem has the intention of arousing the reader really means that the poem's author intended to arouse the reader, the claim that an action has a particular intention means that the *agent* had this

intention. By contrast, people ascribe intentionality to *actions.* What they judge as intentional (or unintentional) are concrete actions, already performed or in the process of being performed.[2]

Second, intentions and intentionality differ in their conditions of ascription. Whether people ascribe an intention to an agent depends on the beliefs and desires they ascribe to that agent—minimally, the agent's desire for an outcome and his belief that the intended action will lead to that outcome. By contrast, people ascribe intentionality based on more stringent criteria that include whether the agent had the intention to act that way, whether she had the skill to do it, and whether she had awareness, at the time of acting, that so acting fulfills her intention (Malle & Knobe, 1997a; Studies 3 and 4).

Third, ascriptions of intention and intentionality differ in their social function. People ascribe intentions to others, and express their own intentions, to describe or coordinate personal and joint future action (e.g., Bratman, 1987; 1993). Even linguistic evidence supports this close link between intention and future action. Having examined 70 different languages, Bybee, Perkins, and Pagliuca (1994, ch. 7) concluded that all linguistic future forms have developed out of forms that signal the speaker's intention. Intentions, then, clearly have a future-directed and coordinating function. By contrast, people ascribe intentionality in order to assess the meaning and evaluate the social worth of a completed (past or present) action.

To summarize, in people's folk theory of behavior, intentions are an agent's mental states directed towards future action whereas intentionality characterizes an action as based on an intention and performed with skill and awareness. This distinction leads to several interesting and previously unexplored questions regarding people's judgments of praise and blame.

Questions Examined in the Present Studies

First, what underlies the intensification effect of intentional action? Is this effect merely due to the presence (vs. absence) of an *intention* that matches the produced action or is it due to the action's *intentionality* (i.e., its performance with awareness and skill)? We try to answer this question in Study 1.

Second, given that evaluations of *actions* are susceptible to the intensification effect of intentionality, are evaluations of *intentions* susceptible to this effect as well? That is, are intentions evaluated differently depending on whether their corresponding action is performed intentionally vs. accidentally? This question is also addressed in Study 1.

Third, is there a general difference between evaluations of a bare intention (an unfulfilled plan) and evaluations of the action that carries out that intention? That is, do people "discount" (i.e., evaluate less extremely) bare intentions relative to their corresponding actions, as some

---

[2] Lawyers and philosophers sometimes ascribe intentionality to an agent in the derivative sense of having the abstract capacity to reason and to act intentionally (a sense that presupposes the ascription of intentionality to actions).

religious and ethical systems suggest (e.g., Judaism or utilitarianism)? We try to answer this question in Study 2.

Fourth, are there asymmetries between praise and blame in any of the above patterns of evaluation? For example, does the intensification effect of intentionality hold more strongly for blame than for praise? Such evaluative asymmetries might parallel positive-negative asymmetries that psychology has uncovered in other domains (Skowronski & Carlston, 1989; Malle & Horowitz, 1995; Pratto & John, 1991; Taylor, 1991). This question is addressed in both studies.

Study 1

In this study we examine what underlies the well-known effect that praise and blame are intensified for intentional as compared to accidental actions. This intensification could be due to two factors: the mere presence of an intention or the action's intentional performance. According to the first analysis, expressed in legal scholarship and the law (Duff, 1990; Hart, 1968; Kenny, 1973; Marshall, 1968), the critical difference between accidental and intentional action is the mere *presence of an intention* in the intentional case. An analysis of people's folk concept of intentionality, by contrast, suggests that an intentional action is not sufficiently defined by the mere presence of an intention (Malle & Knobe, 1997a; cf. Mele & Moser, 1992). Even though the presence of an intention invites more blame or praise than the absence of an intention, intention is only one of three necessary conditions that people use to ascribe intentionality, with the other two being awareness and skill. That is, an agent may have the intention to perform action A but still *accidentally* perform A (without awareness or skill). In this case people say "She intended to A" but not "She intentionally A'ed."[3] According to people's folk concept of intentionality, then, the critical difference between accidental and intentional action is the *intentional performance* of the action— i.e., the presence not only of an intention but also of skill and awareness.

To distinguish these two analyses we shall call the first the *legal account* and the second the *folk-conceptual account*. Both accounts predict intensification of praise or blame when an intention is present, compared to when an intention is absent. For example, if an agent had the intention to perform an action and then accidentally performs it, this should elicit more praise or blame than if the agent did not have such an intention to perform the action and accidentally performs it. In this case, intentionality does not come into play (only intention does), so the two accounts make the same predictions. Similarly, for canonical intentional actions (in which

[3] A famous philosophical example of an agent's action with an intention but without *awareness* is Chisholm's (1966) murderer : He intends to kill his uncle in order to inherit his money but then accidentally runs his uncle over with his car (because he did not see a man, who was his uncle, cross the street). Examples of accidentally acting with an intention but without *skill* abound in the amateur sports world: A novice may intend to hit a triple 20 in darts or a hole-in-one in golf or a three-pointer in basketball, but if the person does so, most would say it was not intentional but due to luck (cf. O'Shaugnessy, 1980). An empirical test of the darts case was validated in Malle & Knobe (1997a), where people did not begin to ascribe intentionality to the novice who hits the triple 20 unless he hit it twice in a row (thus demonstrating skill).

intention and intentionality co-occur), both accounts predict intensification—the legal account because of the presence of an intention and the folk-conceptual account because of the presence of intentionality. However, there are situations in which the two accounts make differential predictions. One such situation is the comparison between (a) an agent who has an intention to perform *A* and then *intentionally* performs *A* and (b) an agent who has an intention to perform *A* and then *accidentally* performs *A.* According to the legal account, only the presence of an intention matters, so people should evaluate the two actions the same way as long as the agent's intentions are identical in each case, even though one action was performed accidentally and the other one intentionally. According to the folk-conceptual account, however, intentionality matters over and above intention, so people should evaluate the accidental performance of the action much less extremely than the intentional performance of that action, even though the agents had identical intentions.

In addition to comparing the legal and the folk-conceptual accounts of the intensification effect, we explore two further issues in Study 1. For one, we examine whether the intensification effect holds only for evaluations of an agent's *action* or also for evaluations of the agent's *intention*. If Malle and Knobe's (1997a) model of the folk concept of intentionality is correct, people distinguish sharply between intentions and actions. Accordingly, praise and blame for actions is susceptible to the distinction between intentional vs. accidental performance, whereas praise and blame for intentions is independent of this performance distinction. An alternative hypothesis is that people's pattern of evaluating intentions follows that of evaluating actions. This could be because people do not see an intention as a separate mental state but rather as a proxy for action (Anscombe, 1957, §4, §26; Ryle, 1949; Wittgenstein, 1953, §644, §659) or because, even thought they distinguish between actions and intentions, they use the directly observable actions as the basis for evaluating intentions. Either way, evaluations of intentions would be equivalent to evaluations of their corresponding actions, such that if an action is intentional and its evaluation is intensified, the evaluation of its underlying intention would be equally intensified. We refer to this alternative view as the *action-intention equivalence thesis.*

The final issue to be explored in Study 1 concerns potential asymmetries between praise and blame. To this end we present two different vignettes in which the respective actions are identical but the motivation for the agent's intention is either benevolent (inviting praise) or malevolent (inviting blame). This way we can determine whether the intensification of intentionality holds as strongly for praise as for blame.

To investigate these three issues, we devised a study that manipulated the valence of an action as well as its accidental vs. intentional performance, and we assessed people's evaluations (praise or blame) for the agent's intention to act as well as for the agent's performing that action. We held the content of the agent's intention constant across all conditions and measured subjects' perceived strength of the agent's intention to control for its potential mediating influence.

Method

Participants. In a mass-testing session of about 200 introductory psychology students, 170 completed the relevant one-page questionnaire (which was part of a larger packet of unrelated questionnaires) and received partial credit towards their course requirement. No demographic data were collected, but comparable samples from the same student population have had a median age of 18 and contained 65% females and 24% ethnic minorities.

Design. The main factors of the design were valence (positive or negative behavior content), intentionality (accidental or intentional performance of action), and object of evaluation (intention or action); only the latter was a within-subjects factor. We counterbalanced the order of questions asked: Some participants evaluated the agent's intention immediately after reading about it (before they read about the completed action) and evaluated the action afterwards, whereas others evaluated both action and intention after they had read the entire story. Within the latter group, we also counterbalanced the order of rating the action and the intention; within the former group, we counterbalanced whether people evaluated the action before or after rating the *strength* of the initial intention.

Material. The story [and its negative counterpart] described an "employee's phone call."[4] Joan Ellen Edmonds, the featured agent, is a clerk at a local company. Her company hired a new clerk, Jonathan Baite, and Edmonds appreciates him quite a bit [strongly dislikes him]. Edmonds heard that Baite loves to get phone calls [absolutely hates to get phone calls at home], so she joyfully [gleefully] decides to give him an appreciation call the next morning when she gets to work [a "wake-up call" next morning when she gets to work at 6 a.m.]. Pretesting showed that blame ratings for the negative story were equal in extremity to praise ratings for the positive story.

The next part of the story manipulated whether the action was performed accidentally or intentionally. When Edmonds arrives at her office the subsequent morning, she also remembers that she wanted to call her mother. She dials that number, but—in the intentional case—nobody answers. Then she calls Baite, who is delighted (extremely annoyed). In the accidental case, she dials that number, but—due to a central switchboard error—she ends up reaching Baite, who is delighted (extremely annoyed).

All subjects were asked the manipulation check question, "Did Edmonds make this call to Baite intentionally?" In addition, all were asked (in varying order), "Does Edmonds deserve blame or praise for *calling* Baite?" Subjects checked a box for either blame or praise and filled in the corresponding word into an empty space embedded in the following two critical questions: "How much _____ [blame/praise] does she deserve for *calling* Baite?" and "How much _____ [blame/praise] does she deserve for initially *intending* to call Baite?" (both 0-7 scales, labeled "a

[4] After this study was conducted we discovered that Mele (1992, p. 151, n. 2) had discussed a similar thought experiment.

little" between the digits 1 and 2 and "a lot" between the digits 6 and 7). In addition, all subjects were asked "How strong was Edmonds' initial intention to call Baite?" (0-10 scale, labeled "very weak" between the digits 0 and 2 and "very strong" between the digits 8 and 10).

Results

Ninety-four percent of subjects passed the manipulation check. Out of 83 in the intentional condition, 80 (96%) rated the behavior as intentional; out of 87 in the accidental condition, 79 (91%) rated the behavior as not intentional. There were no differences in the passing rate either between the positive and negative story or between order conditions. Those 11 subjects who failed the manipulation check were excluded, and 11 more subjects had incomplete data. Analyses were performed on the remaining 148 subjects. No reliable order effects were found, and the results show three noteworthy patterns.

First, in their evaluations of actions (Figure 1), people sharply distinguished between accidental performance (M = 2.0) and intentional performance (M = 4.8), yielding an effect size of d = 1.48 (95% CI: 1.18 to 1.78)[5]; $F(1,146) = 81.9$, $p < .001$. People made this distinction even though the identical intention was present in both conditions and their ratings of strength of intention did not differ ($F < 1$). This result runs counter to the predictions of the legal account and supports the folk-conceptual account: People intensify blame or praise when a behavior is *performed intentionally* (rather than accidentally), even if the appropriate intention is present in both cases.

Insert Figure 1 here

Second, people's evaluations of the agent's *intention* (Figure 1) were identical whether that intention was fulfilled by an intentionally performed action (M = 4.8) or by an accidentally performed action (M = 4.8). This result contradicts the action-intention equivalence thesis and supports the claim that people distinguish between intention and action in their judgments of praise and blame.

Third, the power of intentionality to intensify evaluations of actions appears to be greater for the blaming of negative actions than for the praising of positive actions (Figure 2). Blame for an action performed intentionally (M = 5.7) was intensified by a factor of 3.8 over blame for the same action performed accidentally (M = 1.5), whereas praise for an action performed intentionally (M = 4.1) was intensified only by a factor of 1.6 over praise for the same action performed accidentally (M = 2.5), $F(1, 144) = 19.3$, $p < .0001$. In other words, the effect size for

---

[5] We report a sample-based estimate of effect size d, computed as $\frac{M_1 - M_2}{\sqrt{MS_e}}$, where MS_e is the error term from the corresponding ANOVA (Hedges, 1981). The reported confidence interval for d is based on $s_e = \sqrt{\frac{MS_e}{N}}$.

blame intensification was $d = 2.32$ (95% CI: 1.91 to 2.73), whereas the effect size for praise intensification was $d = 0.88$ (95% CI: 0.48 to 1.28). Controlling for the rated strength of intentions did not alter these results.

We replicated these findings in another sample of 94 undergraduate students. Within intentions, there was again no effect of intentionality ($F < 1$). Within actions, however, we replicated the intensifying effect of intentionality, $F(1,90) = 30.9$, $p < .0001$, $d = 1.12$ (95% CI: 0.84 to 1.4). This intensifying effect was larger for negative actions ($d = 1.5$) than for positive actions ($d = 0.8$), even though the corresponding interaction term did not reach the traditional significance level ($p = .11$).[6]

Discussion

The present findings suggest that people use a refined folk theory of behavior in their assignments of praise and blame. They distinguish between an agent's intention to do something and the intentionality of doing it; they distinguish between evaluations of intentions and evaluations of actions; and they intensify blame more than praise as a function of intentionality. We discuss each of these findings in turn.

To test whether people evaluate actions using intentionality information over and above intention information, we held the influence of intention constant. For one, we used the identical intention content ("to call Baite") in the accidental and intentional condition. This was also the content of the corresponding action, so the agent performed either intentionally or accidentally the very action that she had intended to perform. In addition, we measured the perceived strength of intention and found no difference between the accidental and the intentional condition. Consequently, the differential evaluation of the accidental vs. intentional case must have been due to the presence or absence of *intentionality*—specifically, to the component of intentionality that we manipulated: awareness (Malle & Knobe, 1997a). This component refers to the agent's awareness that she is fulfilling her intention with this particular action (cf. Searle, 1983), which provides the "right link" between intention and action required for intentionality. In sum, information about intentionality (above and beyond intention) can substantially alter people's evaluations of action, favoring the folk-conceptual over the legal account. The distinction between intention and intentionality (Malle & Knobe, 1997a) is therefore not only of conceptual interest but has direct implications for how people assign praise and blame to human action.

Second, we have seen that people evaluate intentions differently from actions (as the folk-conceptual account predicts) in that only evaluations of actions are susceptible to the intentional-accidental distinction. This distinction applies strictly to the performance of an action, not to the

---

[6] Effect sizes were smaller in this replication sample, probably because we unwisely used bipolar scales (–5 to –1 for blame and +1 to +5 for praise) with a neutral point at 0. In particular, the presence of a noncommittal neutral point (chosen by 35% of subjects in their action evaluations) may have dampened the effects.

intention of performing it. Despite this difference between evaluations of actions and intentions, the two are closely related because they are based on the same content (e.g., the intention to call and the action of calling) and on the same motivating reasons (e.g., to annoy Baite in the positive case and to delight Baite in the negative cases). Our study further suggests that these two factors—intention/action content and motivating reasons—are independent determinants of praise and blame. Even though the content of the two intentions was held constant (to call Baite), people evaluated them differentially because one was based on a positive motive (to delight Baite) and the other was based on a negative motive (to annoy Baite). Of course, intentions can differ directly in content and be either blameworthy (e.g., Penny intends to steal money from her father's cash register) or praiseworthy (e.g., John intends to help a neighbor fix his roof). But even in these cases the agent's motivation for the intention will influence people's evaluations (consider that Penny intends to steal money so that she can leave her abusive father). How exactly people "compute" praise and blame on the basis of all these parameters—intention content, motivating reasons, and in the case of action, accidental vs. intentional performance—has not been analyzed systematically and is worthy of future research.

Third, we found an asymmetry between people's use of intentionality for praising vs. blaming an action. This asymmetry held even when we controlled for the perceived strength of intention. Apparently, when people evaluate a negative action, they somehow take intentionality more into account than when they evaluate a positive action. That is, for negative actions they are particularly sensitive to the manner in which the action is performed. This tendency would explain why people assign more blame to negative acts than omissions (equated for intention and outcome) but assign equal praise to positive acts and omissions (Spranca, Minsk, & Baron, 1991). Omissions typically lack actions' sharp boundaries (onset, offset, and bodily movement), creating some doubt about the agent's awareness and full intentional "performance" of the omission (cf. Kordes-de Vaal, 1996). If perceivers are more sensitive to these features of intentionality in the case of negative actions, they would be inclined to discount their blame for negative omissions but not necessarily their praise for positive omissions, as the literature suggests.

Why might people have a heightened sensitivity to the agent's manner of performing negative actions? Two important processes may operate in parallel, one on the intrapersonal level, the other on the cultural level. On the intrapersonal level, the costs of mistaking an intentional negative action as accidental are high (e.g., one is vulnerable to a surprise attack); similarly, the costs of falsely blaming an accidental negative action as if it were intentional are high as well (e.g., one would be seen as unjust). As a result, people need to distinguish carefully between intentional and accidental negative actions. By contrast, the costs of mistaking an intentional positive action as accidental are low (one can still consume the positive outcome), and the costs of falsely praising an accidental positive action as if it were intentional are also low, so people can afford to be somewhat sloppy in distinguishing the two. On the cultural level (at

least in Western cultures, broadly conceived), the distinction between intentional and accidental negative actions has been continually reinforced. One of the earliest documentations of the intentional-accidental distinction refers to negative actions, namely, the Old Testament's distinction between intentional murder and accidental killing (Telushkin, 1994, ch. 58; see Kenny, 1973). The concept of intentionality entered the English law in the 12th century when religious concepts of free will and sin, influential in the Roman Law and Canon law, led to differentiations of punishment—e.g., more punishment for freely chosen crimes (Marshall, 1968). Religion and the law, then, have taught people how to differentiate blame depending on intentionality, but few cultural institutions teach people how to differentiate praise depending on intentionality. One of the few historic institutions of praise, the "ars laudandi" in Renaissance Rome, had orators speak of God's and humans' laudable deeds without emphasizing their (assumed) intentionality (O'Malley, 1979). The same holds true for modern equivalents such as Nobel prizes, Olympic Games, and Oscars, where successes are assumed to be intentional and never discounted as "luck" (cf. the "hot hand" phenomenon, Tversky & Gilovich, 1989).

Study 2

In the first study and its replication, people praised or blamed intentions about as strongly as their corresponding (intentionally performed) actions. This may have been because the vignettes portrayed the respective intentions as having a high degree of commitment: Each intention was based on clearly formulated motives; the agent was described as actively deciding to carry it out; and the intention was in fact fulfilled. In real life, intentions are more often described (or expressed) with less than full commitment, without mention of their motives, and without indications yet of being carried out (e.g., "She plans to come next summer"; "I'll return your papers next week"). Some authors have labeled such yet unfulfilled prior intentions "bare intentions" (e.g., Anscombe, 1957; Kenny, 1973). Study 2 tests the hypothesis that people assign less extreme blame or praise to bare intentions than to their fulfilling actions. Such a phenomenon of "discounting" intentions may occur for at least two reasons.

First, bare intentions are usually removed in time from the intended action and are subject to psychological and situational uncertainties (e.g., change of heart, emergence of obstacles). Hence they are indicators with an imperfect predictive relationship to their fulfilling actions, and imperfect indicators are likely to be discounted relative to their criterion.

Second, when bare intentions are formed in interpersonal contexts, they differ in social meaning and relevance from actions. Intentions are merely symbolic actions. As plans for future action they rarely have (physical) consequences. They can have powerful psychological consequences (e.g. by inducing another person to prevent the plan), but to have those psychological consequences, intentions must be believed. This belief comes in various degrees, leaving room for discounting in the case of weak belief. Actions, by contrast, do not require belief; they speak through their tangible physical and psychological consequences. Thus, bare

intentions may on average be discounted (i.e., evaluated less extremely) relative to their corresponding actions.

In addition to this general discounting pattern, there may also be an asymmetry of discounting between blame and praise. That is, people's blame for negative intentions may show less discounting than does their praise for positive intentions. Such an asymmetry may be due to three processes. First, because of the deviance and infrequency of negative actions, social perceivers may consider negative intentions as predictively more diagnostic of their corresponding (fulfilling) actions than they consider positive intentions as diagnostic of *their* corresponding actions. As a result, people will be more skeptical about positive intentions and limit their praise for them (as they do, for example, for New-Years resolutions). Second, negative intentions may be seen as more diagnostic of the agent's underlying negative character than positive intentions are seen as diagnostic of the agent's underlying positive character (cf. Jones & Davis, 1965; Reeder & Brewer, 1979; Skowronski & Carlston, 1987). As a result, blame for positive intentions would be magnified because it is based on the agent's inferred negative character above and beyond the agent's manifest intention. Third, people may be strategically biased towards attending to and blaming negative intentions more than towards attending to and praising positive intentions because they try to avoid the high costs of overlooking negative intentions that later are fulfilled and see few costs in overlooking positive intentions.

In sum, our second study explores (a) the general discounting effect of bare intentions relative to completed actions across a variety of behaviors and (b) a possible asymmetry in people's blame and praise for bare intentions: that they may, on average, discount negative intentions less than positive intentions.

Method

In order to select balanced stimulus behaviors that had similar characteristics within the negative and the positive range, we conducted a pretest.

Pretest. We first constructed 44 behavior contents, some taken from Fuhrman, Bodenhausen, and Lichtenstein (1989), others invented by friends and colleagues. Each behavior content was formulated both as an intention (e.g., "[Name] intends to help a neighbor fix his roof") and as an action ("[Name] helped a neighbor fix his roof"). Undergraduate students (N = 87) rated either all 44 action formulations or all 44 intention formulations for one (and only one) control variable (e.g., desirability of action, length of action, frequency of intention, out-of-the-ordinariness of intention; see Table 1). Thus the pretest material consisted of a total of 11 forms, one for each control variable, and 7 to 9 subjects completed each form. We computed reliabilities among raters within each control variable and excluded seriously unreliable raters. (The reliabilities along with number of raters, before and after any exclusions, are summarized in Table 1.) Then we calculated the 11 control variable scores (aggregated across reliable raters) for each of the 44 behavior contents to proceed with the selection of appropriate test items.

Insert Table 1 here

Selecting Test Statements. We rank-ordered the 44 pretested behavior statements from the most negative to the most positive (as defined by the action desirability score), with all control scale ratings tabulated for each statement. Then we selected pairs of behaviors from either end (one positive, one negative) that were matched for the extremity of their desirability score (e.g., –3.6 and +3.6). When more than one (positive or negative) behavior had such a matching score (e.g., two behaviors with -3.6 desirability) and thus competed for selection, we used the competing behaviors' control scale scores as a decision criterion. Some of the control scales showed asymmetries between positive and negative behaviors as a group. (For example, positive behaviors had a higher score on length and frequency, whereas negative behaviors had a higher score on out-of-the-ordinariness and difficulty.) Each time we found competing behaviors, we discarded the one that contributed more strongly to the asymmetries. As a result, we usually retained the behavior that was most similar to the matched positive behavior. This way, we were able to weaken or eliminate some of the control scale asymmetries while selecting positive and negative behaviors that were as closely matched as possible. Continuing this selection procedure, we arrived at a matched set of 12 negative and 12 positive statements. These 24 distinct statements were formulated (a) as 24 completed actions and (b) as 24 intentions, making at total of 48 statements. The 24 intentions were identical to the actions except that each action phrase, "[Name] [past tense verb] [behavior content]," was reformulated into an intention phrase, "[Name] [intends to] [verb] [behavior content]."

Creating Test Forms. In order to minimize demand characteristics we wanted to present each participant with either the action formulation or the intention formulation of a given behavior content (e.g., making vs. intending to make a prank phone call), but not both. At the same time, we wanted each participant to rate all 24 distinct behavior contents. We therefore created forms containing 24 distinct statements, half of which were the intention formulations of 12 behavior contents and half of which were action formulations of the remaining 12 behavior contents. From these "base" forms we created "yoked" forms that were identical to the base forms except that the 12 intention statements were transformed into their corresponding action statements and the 12 action statements were transformed into their corresponding intention statements.

In order to maintain participants' attention and to facilitate the incidental memory task (remembering "whether each behavior was completed or intended"), we presented the 24 statements contained in each form in two trials of 12 statements each (with a 30-second break between trials). Within each trial, we selected twelve individual statements with the following characteristics: (a) Six of them were intentions; six were actions. (b) Three of the actions were positive, and the other three were negative with matched extremity; likewise, three of the intentions were positive, and the other three were negative with matched extremity. This created

a total of four "triplets." (c) Each of these triplets was composed of statements that together represented the broad range of desirability—that is, each triplet comprised one statement from the most extreme terzile of desirability, one statement from the intermediate terzile of desirability, and one statement from the least extreme terzile of desirability (e.g., one positive-action triplet comprised the first, fifth-, and ninth-most positive action).[7]

As a result of this procedure, each form contained all 24 distinct statements, but 12 of them were formulated as actions and the other 12 as intentions. To create each yoked form, we simply reformulated all intentions in the base form as actions and all actions as intentions, while leaving everything else identical. In this way we created a total of eight base forms (differing in the choice of triplets from the desirability range) plus eight yoked ones, for a total of 16 forms. Within each form, the first randomized trial of 12 statements was paired up with a list of 12 names (6 female, 6 male), and the second trial was paired up with a list of 12 different names. The order of these names was mostly held constant across forms, with slight modifications to ensure that each statement was paired up with a female name about as frequently as with a male name.

Subjects and Procedure. Subjects were 78 undergraduate students (60% female) enrolled in an introductory psychology course. Their ethnic composition was 74% White, 18% Asian American, 6% Hispanic, and 1% African American. All but six were native English speakers. Subjects were run in groups of 8-12. Each subject received a packet that introduced, after instructions, the first set of 12 behavior statements (allowing them 2 minutes to memorize whether a behavior was described as intended or performed), then asked people to rate each statement on a blame-praise scale (ranging from -5 to +5), and finally tested people's memory for whether each behavior had been described as intended or performed. After a 30-second break, subjects went through the same steps with the second set of 12 behavior contents.

Results

We first inspected the rating distributions for each of the 24 behavior contents to identify outliers. We defined outlier data points as extreme ratings (4 or 5) with a valence opposite to the

[7] This procedure is best illustrated by describing the steps for one particular form, namely, Form 1. We began with the selection of 12 statements for the first trial. We first selected a triplet of positive actions from the 12 positive action statements (rank-ordered by desirability and labeled PA-1 to PA-12) by choosing one statement from each terzile (PA-1 from the first terzile, PA-5 from the second terzile, and PA-9 from the third terzile). Next, we found the extremity-matched negative actions (NA-1, NA-5, NA-9), increasing the selection to 6. Then we selected a triplet of positive intentions from the 12 positive intention statements (rank-ordered by desirability and labeled PI-1 to PI-12) by selecting one statement from each terzile (PI-2 from the first terzile, PI-6 from the second terzile, and PI-10 from the third terzile). Finally, we found the matched negative intentions (NI-2, NI-6, NI-10), increasing the total selection to 12. (Note that the selected intention statements are different in content from the selected action statements.) We then randomized the order of presentation for these 12 statements. Then we turned to the second trial for which we selected those statements that were not already selected in the first trial: PA-3, PA-7, PA-11, NA-3, NA-7, NA-11, and PI-4, PI-8, PI-12, NI-4, NI-8, NI-12. We also randomized the order of presentation for these 12 statements.

rest of the sample (and to the pretest data), such as a +5 praise for "sold cocaine to his teenage cousin." The 1872 data points contained 21 (1%) such outliers. One subject contributed seven of them, another four, and both subjects' entire data were excluded from subsequent analyses. The remaining 10 outlier points were distributed across 9 different subjects and 7 different behaviors and were probably the result of accidental mismarkings or idiosyncratic construals of particular behaviors. These 10 data points (but not the subjects' entire data) were excluded from analyses.

People's accuracy of remembering whether a given behavior was intended or performed was better than 95% (23/24 correct on average). Accuracy was slightly greater for actions (96.5%) than intentions (94%), $F(1,75) = 3.92$, $p = .05$. There was also a modest interaction such that negative intentions and positive actions yielded greater accuracy than negative actions and positive intentions, $F(1,75) = 3.54$, $p = .06$.

The two hypotheses—discounted evaluations of intentions relative to actions and greater discounting for positive intentions than for negative intentions—were tested in several ways, the first one with subjects as units of analysis. Each subject contributed four scores: praise/blame for the average of (a) six negative actions, (b) six negative intentions, (c) six positive actions, and (d) six positive intentions. These scores were entered into a 2 (positive vs. negative valence) × 2 (intention vs. action) ANOVA with repeated measures on both factors and $N = 76$ replications per cell. The negative signs of blame scores were removed at this point so that the analysis would pick up extremity differences between positive and negative evaluations rather than trivial valence differences. Overall, evaluations of intentions ($M = 2.4$) were discounted relative to actions ($M = 2.8$), $d = 0.48$ (95% CI: 0.29 to 0.67), yielding $F(1,75) = 24.7$, $p < .001$. In addition, when we examined discounted blame for negative intentions ($M = 2.6$) relative to negative actions ($M = 2.9$), the effect size was $d = 0.36$ (95% CI: 0.17 to 0.55). By contrast, when we examined discounted praise for positive intentions ($M = 2.1$) relative to positive actions ($M = 2.8$), the effect size was more than twice as large, $d = 0.84$ (95% CI: 0.65 to 1.03). The corresponding interaction yielded $F(1,75) = 7.7$, $p < .01$. Using subjects' memory accuracy as covariates in this analysis left all effects unaltered.

This analysis was repeated with stricter scoring procedures as well. Only those behaviors were included in individual subjects' scores that had substantial ratings of blame or praise within the entire sample (≥ 2.0, with 0 as the neutral point and 5.0 as the most extreme rating possible), yielding scores based on 10 clearly positive and 10 clearly negative behaviors (out of 24 total). Evaluations of intentions ($M = 2.7$) were discounted relative to actions ($M = 3.3$), $d = 0.60$ (95% CI: 0.38 to 0.82), $F(1,75) = 26.0$, $p < .001$. In addition, the discounting effect size for blame was $d = 0.40$ (95% CI: 0.18 to 0.62),whereas the discounting effect size for praise was twice as large at $d = 0.80$ (95% CI: 0.58 to 1.02). The corresponding interaction yielded (1,75) = 7.3, $p < .01$.

Insert Table 2 here

Next we used individual behavior contents (N_b = 24) as units of analysis. Table 2 displays each behavior content's action score and intention score, averaged across 35 to 39 subjects. To test our hypotheses we conducted a 2 (positive, negative) × 2 (intention, action) ANOVA on these highly aggregated scores. Overall, intentions (M = 2.4) were discounted relative to actions (M = 2.8), d = 1.28 (95% CI: 1.13 to 1.43), $F(1,22) = 19.5$, $p < .001$. In addition, there was more than twice as much praise discounting among the positive behavior contents (d = 1.82, 95% CI: 1.67 to 1.97) than there was blame discounting among the negative behavior contents (d = 0.74, 95% CI: 0.59 to 0.89). The corresponding interaction yielded $F(1,22) = 3.45$, $p < .08$. This effect of valence was similar when we restricted the comparison to the 10 strictly positive behavior contents (d = 2.21) and the 10 strictly negative behavior contents (d = 1.02).

We also inspected the pretest control measures to determine whether any of them substantially reduced (and thus explained) the two discounting effects. However, none of these measures (except, of course, social desirability) had correlations with the discounting scores, and none of them altered the above hypothesis tests when entered as covariates into the ANOVA.

Finally, to demonstrate that the blame-praise asymmetry is due to differences in the discounting of *intentions* rather than to base rate variations in the evaluation of *actions*, an ANCOVA analyzed the behavior contents' intention scores as a function of (a) action scores and (b) the dichotomous valence factor representing negative vs. positive behavior contents. This analysis equates positive and negative behavior contents for their extremity in action scores and tests a difference in their intention scores strictly due to valence. This latter difference emerged strongly at d = 1.0, $F(1,21) = 5.3$, $p < .05$.

Discussion

Study 2 examined how people's evaluations of bare (yet unfulfilled) intentions differ from evaluations of completed actions. The results support the hypothesis that people generally discount their praise and blame of intentions relative to their corresponding actions. The results also support the hypothesis that people discount praise for positive intentions more than they discount blame for negative intentions.

How generalizable are these discounting effects? We consider three aspects of generalizability. First, in Study 2 we examined a large number of behavior contents that varied on many dimensions (see list of control variables, Table 1). Across these behaviors we found consistent evidence for a general discounting of intentions and for greater discounting of positive than negative intentions. In addition, we replicated these findings in another study with 36 distinct behavior contents, ranging from extremely kind to extremely unkind, as normed by Fuhrman et al. (1989). A total of 127 subjects rated sets of either 9 or 18 such behaviors, in the same procedure as described above, and the results were essentially the same. (For details, see

<http://www.uoregon.edu/~bfmalle/36beh.html>.) We therefore believe that the results generalize well to a variety of behavior contents.

Second, the results of this study generalize only to bare intentions, not to intentions that immediately precede or coincide with their corresponding actions (see Study 1). We expect that a variant of Study 2 that presented intentions in action (e.g., "[name] *tried to* [behavior content]") would decrease the general discounting effect. Similarly, highlighting the agent's commitment to the intention—e.g., as we did in Study 1, by complementing the intention statements with well-defined descriptions of the agent's motives and labeling that intention a "decision" to do something—should decrease people's discounting.

Third, 65% of subjects showed a consistent intention-discounting effect, and 65% showed the blame-praise asymmetry in discounting (there was no correlation, however, between showing one or the other effect). It seems plausible that individual differences moderate both effects. People may differ in their moral rules and ideologies (Forsythe, 1981), such that some people make a sharper distinction between intentions and actions than others. For example, Christian doctrines consider mere thoughts about sinful behavior already sinful (most radically expressed by middle-age scholar Peter Abelard, see Kenny, 1973), whereas Judaic teachings focus more on action and outcomes than on intention and motivation (e.g., Blank, 1991, ch. 4; Talmud, 1985, 23A+B). Another potentially relevant variable is the social perceiver's level of interpersonal trust vs. suspicion (Kramer, 1994; Parks, Henager, & Scamahorn, 1996), which may influence her confidence in agents' general inclination to carry out their intentions, especially their negative ones.

How can we explain the present findings? Because none of the control variables accounted for the discounting effects in our data, several explanations for the effects are ruled out, such as those that rely on incidental features of the specific intentions or actions themselves (e.g. difficulty, commonness, extremity of desirability). Below we consider explanations, first, of the general discounting effect and, second, of the praise-blame asymmetry.

An account of the *general discounting of intentions* must start with people's folk concept of intentionality, which posits that an intention is the agent's mental state directed towards future action and is a necessary but insufficient condition for the performance of an intentional action (Malle & Knobe, 1997a). Two facts follow from this concept: First, because they are mental states, intentions have less palpable consequences than actions do and proved fewer opportunities for evaluation than actions do. Second, intentions predict actions only imperfectly, so people will tend to be conservative in their evaluations until they see whether the action really occurs or not. This predictive relationship between intentions and actions is imperfect because for an intention to be carried out, several things must fall into place. For one, the intention has to persist, which usually implies that its motivating reasons (desires and beliefs) must persist. But desires change and new information emerges, which can sometimes alter an agent's intention. Furthermore, for an intention to be carried out, several enabling factors need to

be present (e.g., the agent's skill, an opportunity) and inhibitory factors need to be absent or overcome (e.g., interference from other people or physical obstacles).

To account for the *asymmetry in the discounting of positive vs. negative intentions*, we now consider the three explanations introduced earlier and evaluate them in light of the results.

The first explanation relies on negative intentions' greater *predictive diagnosticity* for negative actions—i.e., their greater likelihood of being turned into action. If people assumed such greater diagnosticity of negative than of positive intentions, they would be justified in discounting positive intentions more than negative intentions. However, our pretests showed that people consider negative intentions significantly *less* likely to be fulfilled than positive intentions (see Table 1)—perhaps because they believe that the world generally works against the fulfillment of negative intentions. Thus, the present data do not support the predictive diagnosticity explanation.

The second explanation posits that negative intentions are seen as more diagnostic of the agent's underlying negative character than positive intentions are seen as diagnostic of the agent's underlying positive character. A similar explanation has been used to account for a negativity effect in impression formation according to which people are more prone to infer negative traits from negative actions than positive traits from positive actions (Jones & Davis, 1965; Reeder & Brewer, 1979; Skowronski & Carlston, 1987). However, our pretests do not support this explanation either. One of the control questions was, "How likely is it that the person has intended to perform this behavior before?" If people drew stronger character inferences from negative than positive intentions, they should be more inclined to assume that the agent has intended this behavior before; but the difference between the 12 negative and the 12 positive behaviors was negligible on this control question, and the means went into the opposite direction (see Table 1).

The third explanation assumes that people have a strategic bias towards greater responsiveness to negative intentions. This explanation is best described in a signal-detection framework. The "detection" problem is this: If person P expresses an intention to perform action A, should you "respond" to P's intention? Responding may include paying attention, taking the intention seriously, expressing praise/blame, expecting that P will in fact perform A and thus changing your own behavior in light of P's anticipated action. Because you either respond or not and the action either occurs or not, there are four cases to distinguish. Let a "hit" be that you respond and the action occurs; a "correct rejection" that you do not respond and the action does not occur; a "false alarm" that you respond but the action does not occur; and a "miss" that you do not respond but the action does occur (see Figure 3). The decision to respond or not derives from the benefits of hits minus the costs of false alarms and the costs of misses minus the benefits of correct rejections. We believe that for positive intentions false alarms are more costly than misses, whereas for negative intentions misses are more costly than

false alarms, suggesting a bias towards responding more strongly to negative intentions than to positive intentions. The reasoning goes like this:

If you ignore P's negative intention and it ends up being carried out (a miss) you risk both the negative action (e.g., a surprise attack) and the possible regret that you did not try to prevent it (e.g., through blame or counteraction). If you respond to P's negative intention and it ends up not being carried out (false alarm) you feel relieved because the action did not occur and perhaps proud because your response may have helped prevent P's action. By contrast, if you ignore P's positive intention and it ends up being carried out (a miss), you can enjoy the positive action (e.g., a surprise gift). If you respond to P's positive intention and it ends up not being carried out (false alarm) you feel disappointed (e.g., because of a broken promise). Thus, people may pay close attention to negative intentions, readily blame them, and initiate preventive measures because it would be costly not to, in case the intention leads to action. (If recall accuracy is an indication of attention, then people's greater recall accuracy for negative intentions in Study 2 supports the notion that they pay more attention to negative than to positive intentions.) By contrast, people may be more reluctant to pay close attention to positive intentions, praise them, or rely on them, because it would be costly to do so and then see the intention not carried out. This strategic bias probably develops in direct social interactions, but it may well transfer to mere observations of others' intentions (which was the situation our subjects were in).

Our data are consistent with this signal-detection analysis, but future research must provide more direct tests, such as by manipulating the costs of misses and false alarms.

Summary and Conclusions

Previous explorations of the folk concept of intentionality (Malle & Knobe, 1997a) led to the claim that people distinguish between an agent's intention and the intentionality of an action. The present studies examined the implications of this distinction for people's praise and blame of both actions and intentions.

Before we summarize our findings we should emphasize that both studies focused on praise and blame from a non-interacting observer perspective. This is a common perspective from which people judge social behavior—e.g., as citizens on jury duty, consumers of media coverage, or perceivers of other's interactions. However, people also blame or praise the intentions and actions directed at *them*, and the present findings do not necessarily generalize to such cases. We are convinced that the same conceptual distinctions hold in those cases (e.g., between intention and intentionality) and expect the effects we found for observers to be even stronger for interactants because the costs of, say, overlooking negative intentional actions or missing a negative intention now have direct implications for the interactant's own welfare. In addition, the interactant's emotions, goals, and personality may influence patterns of praise and blame in systematic ways. We hope that future research will examine these regularities of praise and blame for others' intentions and actions directed at oneself.

Now to a summary of our results. Study 1 and its replication demonstrated that what intensifies praise and blame for intentional (vs. accidental) action is the *intentional performance* of the action, not merely the presence of an intention. This finding supports the folk-conceptual account of the intensification effect and speaks against the legal account of this effect. Moreover, such intensification of praise and blame holds only when people evaluate an action, whereas evaluations of an intention appear to be independent of the manner in which the corresponding action is performed. This finding again supports the folk-conceptual model and speaks against the action-intention equivalence thesis. Study 2 asked people to evaluate bare (as yet unfulfilled) intentions in contrast with completed actions. We found that bare intentions were evaluated less extremely, hence discounted, relative to their corresponding actions. This discounting pattern can be explained by the more palpable consequences of actions compared to intentions and by the imperfect predictive relationship between intentions and actions, which encourages people to be conservative in their evaluations until they see whether the action really occurs or not.

Both studies also demonstrated asymmetries between praise and blame. For one, the intensified evaluation of intentional (vs. accidental) action held more strongly for the blaming of negative action. We speculated that two processes may underlie this asymmetry: On the intrapersonal level, mistaking negative intentional actions as accidental and falsely blaming negative accidental actions as if they were intentional are more costly than the corresponding errors for positive actions. On the cultural level, intentionality has historically been more closely linked to blame than to praise, and judgments of intentionality have been primarily used to differentiate degrees of blame, not praise (Marshall, 1968; Kenny, 1973). As this link between intentionality and blame has established itself in modern cultural institutions, people may have been socialized towards greater sensitivity to the intentional-accidental difference when blaming negative actions than when praising positive actions.

A second praise-blame asymmetry emerged in the evaluative discounting of bare intentions relative to actions: People discounted negative intentions less than positive intentions. This effect is best explained by the strategic bias social perceivers have towards "detecting" negative intentions, which is mainly due to the greater costs of ignoring negative intentions in case they turn into action.

In acknowledging people's folk concept of intentionality and especially the distinction between intention and intentionality, the present studies point to novel regularities in people's praise and blame for social behavior. We hope to have shown that careful attention to people's own folk concepts of behavior offers valuable insight into the complexities of social perception and interaction.

References

Anscombe, G. E. M. (1957). Intention. Oxford: Basil Blackwell.

Aristotle (1892). *The Nicomachean ethics* (J. E. C. Welldon, Trans.). London: Macmillan. (Originally published around 330 B. C.)

Astington, J. W. (1991). Intention in the child's theory of mind. In D. Frye & C. Moore (Eds.), *Children's theories of mind: Mental states and social understanding*. Hillsdale, NJ: Erlbaum.

Blank, W. (1991). *Torah, Tarot, and Tantra: A guide to Jewish spiritual growth*. Sigo Press.

Brand, M. (1984). *Intending and acting: Toward a naturalized action theory*. Cambridge, MA: MIT Press.

Bratman, M. E. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Bratman, M. E. (1993). Shared intention. *Ethics*, *104*, 97-113.

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

Bybee, J., Perkins, R. & Pagliuca, W. (1994). *The evolution of grammar*. Chicago: University of Chicago Press.

Chisholm, R. M. (1966). Freedom and action. In K. Lehrer (Ed.), *Freedom and determinism*. New York: Random House.

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

Duff, R. A. (1990). *Intention, agency and criminal liability*. Oxford: Basil Blackwell.

Fiske, S. T. (1989). Examining the role of intent: Toward understanding its role in stereotyping and prejudice. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought: Limits of awareness, intention, and control* (pp. 253-283). New York: Guilford.

Fleming, J. H., & Darley, J. M. (1989). Perceiving choice and constraint: The effects of contextual and behavioral cues on attitude attribution. *Journal of Personality and Social Psychology, 56*, 27-40.

Forsyth, D. R . (1981). Moral judgment: The influence of ethical ideology. *Personality and Social Psychology Bulletin*, *7*, 218-223.

Fuhrman, R. W., Bodenhausen, G. V., & Lichtenstein, M. (1989). On the trait implications of social behaviors: Kindness, intelligence, goodness, and normality ratings for 400 behavior statements. *Behavior Research Methods, Instruments, and Computers*, *21*, 587-597.

Tversky, A., & Gilovich, T. (1989). The cold facts about the "hot hand" in basketball. *Chance*, *2*, 16-21.

Gopnik, A., & Meltzoff, A. (1996). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Hart, H. L. A. (1968). *Punishment and responsibility. Essays in the philosophy of law*. New York: Oxford University Press.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107-128.

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

Hogue, T. E., & Pebbles, J. (1997). The influence of remorse, intent and attitudes toward sex offenders on judgments of a rapist. *Psychology, Crime, & Law*, *3*, 249-259.

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. 371-388). Hillsdale, NJ: Erlbaum.

Kenny, A. (1973). The history of intention in ethics. In A. Kenny, *The anatomy of the soul. Historical essays in the philosophy of mind* (pp. 129-146). Oxford: Basil Blackwell.

Kleinke, C. L., Wallis, R., & Stadler, K. (1992). Evaluation of a rapist as a function of expressed intent and remorse. *The Journal of Social Psychology*, *132*, 525-537.

Kordes-de Vaal, J. H. (1996). Intention and the omission bias: Omissions perceived as nondecisions. *Acta Psychologica*, *93*, 161-172.

Kramer, R. M. (1994). The sinister attribution error: Paranoid cognition and collective distrust in organizations. *Motivation and Emotion*, *18*, 199-230.

Malle, B. F. (1997, August). *People's folk theory of behavior*. Paper presented at the annual convention of the Cognitive Science Society, Stanford University.
Malle, B. F. (1998). *How people explain behavior*. Manuscript submitted for publication.
Malle, B. F., & Horowitz, L. M. (1995). The puzzle of negative self-views: An explanation using the schema concept. *Journal of Personality and Social Psychology, 68*, 470-483.
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.
Marshall, J. (1968). *Intention in law and society*. New York: Funk & Wagnalls.
Maselli, M. D., & Altrocchi, J. (1969). Attribution of intent. *Psychological Bulletin, 71*, 445-454.
Mele, A. R., & Moser, P. K. (1994). Intentional action. *Nous, 28*, 39-68.
Meltzoff, A. N. (1995). Understanding the intentions of others; Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 1-16.
Moses, L. J. (1993). Young children's understanding of belief constraints on intention. *Cognitive Development, 8*, 1-25.
Moses, L. J., & Chandler, M. J. (1992). Traveler's guide to children's theories of mind. *Psychological Inquiry, 3*, 286-301.
O'Malley, J. W. (1979). *Praise and blame in Renaissance Rome: Rhetoric, doctrine, and reform in the sacred orators of the papal court, c. 1450-1521*. Durham: Duke University Press.
O'Shaugnessy, B. (1980). *The will*. Cambridge, England: Cambridge University Press.
Ossorio, P. G., & Davis, K. E. (1968). The self, intentionality, and reactions to evaluations of the self. In C. Gordon & K. J. Gergen (Eds.), *The self in social interaction*. New York: Wiley
Parks, C. D., Henager, R. F., & Scamahorn, S. D. (1996). Trust and reactions messages of intent in social dilemmas. *Journal of Conflict Resolution, 40*, 134-151.
Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
Piaget, J. (1932). *The moral judgment of the child*. New York: Harcourt, Brace.
Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380-391.
Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86*, 61-79.
Ryle, G. (1949). *The concept of mind*. Chicago: University of Chicago Press.
Schlenker, B. R., Britt, T. W., Pennington, J., Murphy, R., & Doherty, K. (1994). The triangle model of responsibility. *Psychological Review, 101*, 632-652.
Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge, England: Cambridge University Press.
Shaver, K. G. (1985). *The attribution of blame*. New York: Springer-Verlag.
Shultz, T. R. (1980). Development of the concept of intention. In W. A. Collins (Ed.), *Development of cognition, affects, and social cognition. The Minnesota symposium on child psychology* (Vol. 13, pp. 131-164). Hillsdale, NJ: Erlbaum.
Shultz, T. R. (1988). Assessing intention: A computational model. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 341-367). New York: Cambridge University Press.
Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52*, 689-699.
Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27*, 76-105.
Talmud (1985). *The Babylonian Talmud* (Vol. 13, Nazir). (B. D. Klien, Trans.). London: Soncino Press.
Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*, 67-85.
Telushkin, J., Rabbi (1994). *Jewish wisdom*. New York: William Morrow & Co.
Weiner, B. (1995). *Judgments of responsibility*. New York: The Guilford Press.

Wellman, H. (1990). *Children's theories of mind*. Cambridge, MA: MIT Press.
Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford: Basil Blackwell.

Table 1

Control Variables, Their Reliabilities (Assessed for 44 Behavior Statements in Study 2's Pretest), and Their Average Scores for the 12 Positive and 12 Negative Behaviors Selected for Study 2

Control Variable Questions	α (N)	α after excluding unreliable judges	12 positive behaviors	12 negative behaviors
For Action Formulations:				
How socially good or bad is this behavior?	0.97 (8)		2.3	-2.3
How long does it take to perform this behavior?	0.96 (7)		2.4	2.2
How out-of-the-ordinary is this behavior?	0.89 (8)		3.8	4.6
How frequently do people perform this behavior?	0.83 (8)	0.86 (7)	4.0	3.8
How likely is it that the person has performed this behavior before?	0.87 (9)	0.89 (8)	4.6	4.8
For Intention Formulations:				
How socially good or bad is this intention?	0.97 (8)		1.9	-2.3
How difficult is it to fulfill this intention?	0.91 (9)		3.4	4.1
How out-of-the-ordinary is this intention?	0.89 (7)		3.3	4.8*
How frequently do people form this intention?	0.80 (7)	0.82 (6)	3.5	3.0
How likely is it that the person has intended to perform this behavior before?	0.72 (8)	0.77 (7)	4.2	3.8
How likely is it that the person will NOT fulfill this intention?	0.34 (8)	0.76 (4)	2.8	4.9**

Note: * $p < .05$, ** $p < .001$ for comparing absolute control variable scores between the 12 negative and 12 positive behaviors

Table 2

24 Behavior Contents and Their Corresponding Action Scores and Intention Scores, Ordered by Valence of Action Scores, in Study 2.

Behavior content	M act	SD act	M intend	SD intend
sold cocaine to his teenage cousin	-4.44	1.12	-4.33	1.10
started a false rumor about a work colleague's criminal past	-4.00	1.08	-3.62	1.32
disowned her son after he told her he was gay	-3.92	1.59	-2.97	1.80
stole $50 out of her dad's wallet	-3.87	1.53	-3.65	1.40
built a highly explosive plastic bomb	-3.54	1.95	-3.72	1.43
ripped off her brother when buying a joint present for their parents	-3.32	1.60	-2.87	1.73
ate all the food before anyone else could have any	-2.92	1.46	-2.33	1.03
parked her car illegally on a handicapped spot in front of a store	-2.87	1.61	-2.70	1.33
harassed her neighbor's dog	-2.69	1.45	-2.24	1.34
made a prank phone call	-2.32	1.56	-1.77	1.39
spent all summer on a golf ranch without his wife and children	-0.87	2.43	-0.97	2.44
went swimming outdoors in 30 degree weather	-0.16	2.08	-0.41	1.39
bought a coffee mug with her initials on it	0.59	1.45	0.78	1.40
bought an expensive new heating system for her house	0.81	1.47	0.41	1.16
explained national budgetary policy to a colleague	2.05	1.78	1.08	1.13
replaced three lightbulbs in her parents' house	2.33	1.49	2.00	1.31
loaned her computer to a colleague	2.54	1.52	2.68	1.63
put money in the meter for an unknown driver	2.97	1.61	2.58	1.54
bought food for everyone in the office	3.23	1.39	2.84	1.36
called a distant uncle on his birthday	3.32	1.36	2.23	1.49
sowed a silk blouse for her mother	3.51	1.35	3.05	1.45
let a homeless person spend a night in his apartment	3.76	1.53	1.61	1.44
participated in an effort to clean up a city park	3.79	1.20	2.89	1.61
paid a month's rent for a family threatened to be evicted	4.49	0.98	3.05	1.83

Note: The top ten behavior contents are the "strictly" negative ones, the bottom 10 are the "strictly" positive ones.

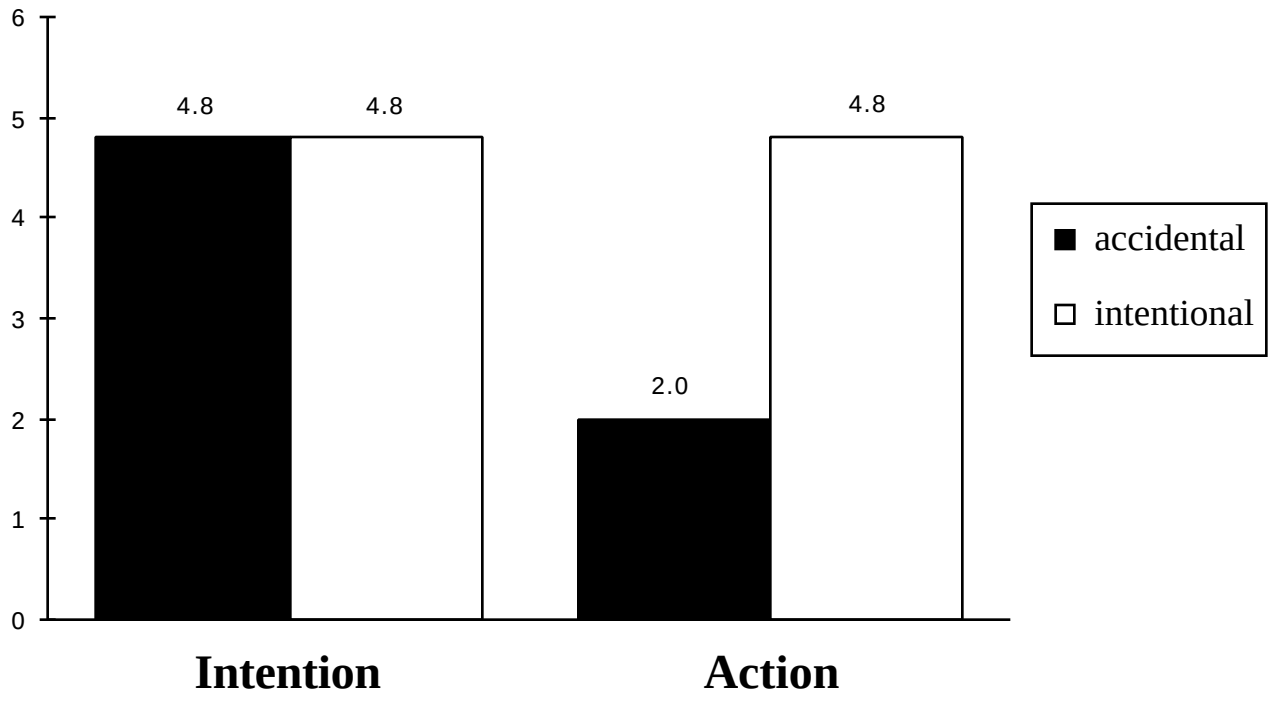

Figure 1.

Extremity of evaluation for intentions and actions
depending on whether the action is performed accidentally or intentionally

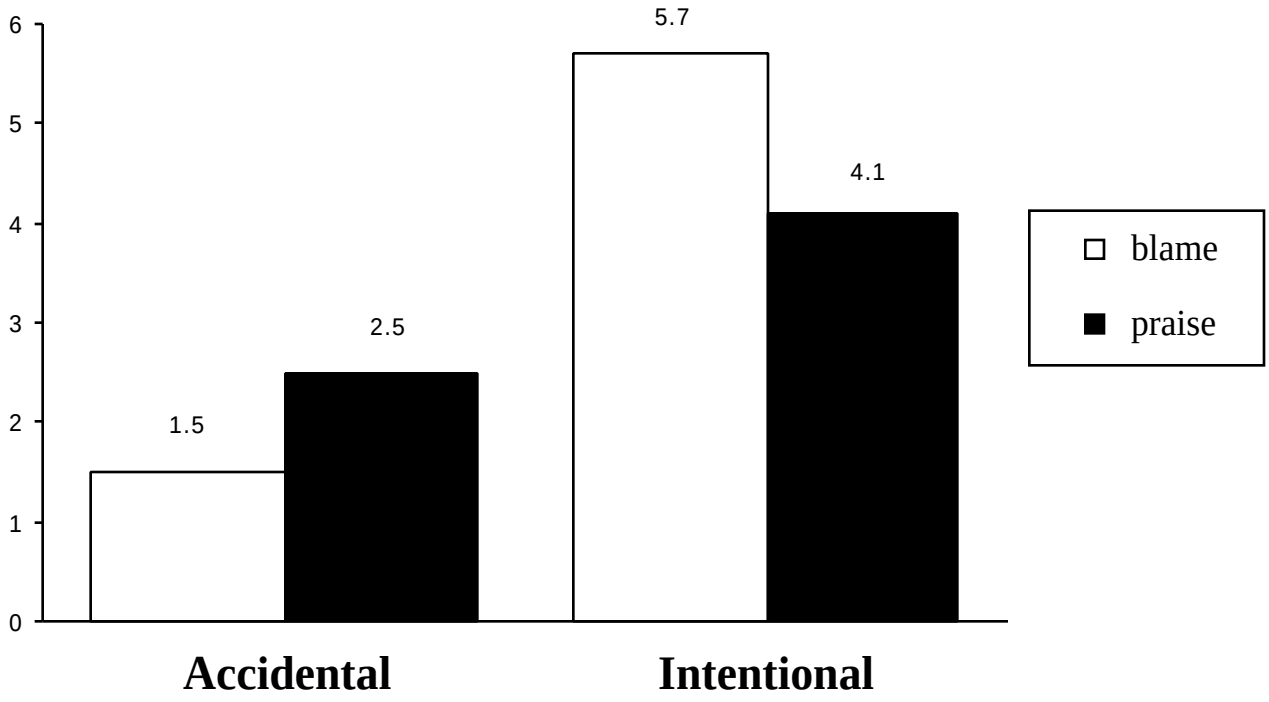

Figure 2.

Blame and praise for actions only
depending on whether they were performed accidentally or intentionally

	Action Occurs	No Action Occurs
Respond to Intention	**hit**	**false alarm**
Don't Respond to Intention	**miss**	**correct rejection**

Figure 3.

Signal-detection analysis of the consequences of responding or not responding to an intention.