CrossMark
click for updates

Perceived intent motivates people to magnify observed harms

Daniel L. Ames and Susan T. Fiske[1]

Department of Psychology, Princeton University, Princeton, NJ 08544

Existing moral psychology research commonly explains certain phenomena in terms of a motivation to blame. However, this motivation is not measured directly, but rather is inferred from other measures, such as participants' judgments of an agent's blameworthiness. The present paper introduces new methods for assessing this theoretically important motivation, using tools drawn from animal-model research. We test these methods in the context of recent "harm-magnification" research, which shows that people often overestimate the damage caused by intentional (versus unintentional) harms. A preliminary experiment exemplifies this work and also rules out an alternative explanation for earlier harm-magnification results. Exp. 1 asks whether intended harm motivates blame or merely demonstrates the actor's intrinsic blameworthiness. Consistent with a motivational interpretation, participants freely chose blaming, condemning, and punishing over other appealing tasks in an intentional-harm condition, compared with an unintentional-harm condition. Exp. 2 also measures motivation but with converging indicators of persistence (effort, rate, and duration) in blaming. In addition to their methodological contribution, these studies also illuminate people's motivational responses to intentional harms. Perceived intent emerges as catalyzing a motivated social cognitive process related to social prediction and control.

intent | motivation | harm | blame | morality

For decades, experimental psychologists, philosophers, legal theorists, and sociologists have worked to understand why and how people blame. These "blame experts" are largely in accord as to what constitutes a blameworthy act. Alicke's (1) description of blameworthiness is representative: "A blameworthy act occurs when an actor intentionally, negligently or recklessly causes foreseen, or foreseeable, harmful consequences without any compelling mitigating or extenuating circumstances." However, although blame experts agree about what kinds of acts elicit blame judgments, they often disagree about the nature of those blame judgments. Arguably the most recurrent disagreement is the extent to which blame judgments are shaped by motivational forces (people's desires and goals) vs. epistemic forces (people's knowledge and reasoning). Whereas few doubt that both motivational and epistemic forces can be involved in the psychology of blame, the relative importance of these psychological forces remains unclear. Some researchers accord motivated cognition a central role in shaping moral judgment (2–6). Others accord much less importance to motivated processes, and correspondingly more weight to inferential reasoning processes and bounded rationality (7, 8), and still others sit somewhere between these two positions (9–11).

From an attributionist perspective, the desire to assign blame for bad actions and their consequences (rather than to accept the notion that bad events arise from disorderly circumstance) stems from a motivation to maintain a sense of control and predictability. According to this model, "Attribution processes are to be understood, not only as a means of providing the individual with a veridical view of his world, but as a means of encouraging and maintaining his effective exercise of control in the world" (12). This idea is distinct from—but related to—the view that moral judg-

ment is largely motivated by the need to evaluate other persons, which allows people to navigate their social environment in more adaptive ways (3, 13, 14). Another possible source of moral motivations is a broader "justice motive" (15, 16), which drives people to correct moral transgressions and to seek equity and fairness. The moral dyads theory (17, 18) offers a different perspective, suggesting that people alter or distort incoming moral information to cohere with a cognitive template that describes people's prototypical views of moral situations. This theory acknowledges both motivated and unmotivated processes that can lead to such distortions, but tends to place greater emphasis on the unmotivated processes (19). The path model of blame (8) delineates a specific cognitive structure (or "path") that people follow to produce blame judgments. Although not denying the possibility of motivated reasoning, this theory nonetheless strongly favors amotivational accounts of phenomena for which other theories posit motivational mechanisms. In contrast, the culpable control model (2) and social intuitionist model (4) place motivation and evaluation at the center of moral judgment. Dual-process theories (9, 10), which parcellate moral judgment into automatic and controlled components, give substantive weight to both motivated moral reasoning and bounded rationality. However, the question of whether a cognitive process is automatic vs. controlled may be orthogonal to the question of whether that process is motivated. The psychology literature shows that both automatic and controlled mental processes can be either motivated or unmotivated (20–23).

Thus, although all societies care about detecting and thwarting harmful behavior, and all agree that holding transgressors accountable through blame is important for maintaining social order, the question of what drives blame (and by extension, how

---

Significance

This paper examines how detecting harmful intent creates downstream consequences for assessing damage, magnifying its cost. If intentional harms seem worse, society may spend more money on them than on objectively more damaging unintentional harms or on naturally occurring harms. Why might this occur? Various psychological theories identify the cause as motivation; however, the presence of this motivation has been inferred indirectly. Drawing on animal-model research, we present more direct evidence for blame motivation, and discuss how it may help to explain the magnification of intentional harms. This approach acknowledges the nonrational biases in damage estimates and potential policy priorities.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

and when blame is exacerbated or mitigated) remains unclear. Researchers with opposing theoretical perspectives often point to the same data as evidence that supports their own theory over the alternative. In many ways, this debate over blame motivation is reminiscent of the "hot vs. cold cognition" controversy that preoccupied experimental psychologists during the 1970s and 1980s (see ref. 24 for review). The apparent intractability of the central issues of this controversy led some researchers to deem the debate unresolvable (25, 26). Ultimately however, considerable progress was made once researchers began to shift their empirical efforts toward defining the mechanisms underlying the phenomena they were studying. This perspective shift opened up more direct avenues for assessing when and how people's "hot cognition" influenced (or did not influence) their perceptions. In a similar vein, the persistence of the current debate over motivated moral cognition stems partly from the fact that, although it is common to explain findings in terms of blame motivation, there have been few if any dedicated attempts to provide direct evidence for the involvement of motivation in blame judgments. Rather, the empirical arguments for and against motivated accounts of blame depend largely on how experimentalists interpret participants' judgments of agent blameworthiness, intent, causation, desire, foresight, foreseeability, and other factors that rational and legal models identify as related to blame. Here, we introduce a new methodological approach for assessing blame motivation, drawn from animal-model research. This approach relies on behavioral measures rather than participants' reports of their own judgments and feelings. The hope is that such methods will be useful to researchers studying moral judgment, and other topics where the role of motivation is unclear.

To help validate these methods, we attempt to apply them to a recent program of work on "harm-magnification" effects. A primary enterprise (perhaps the primary enterprise) of moral psychologists has been to characterize how, when, and why people blame and punish. Thus, the success of many theories have been gauged by their ability to predict how much blame (or punishment) people ascribe across different situations. Although several theories have been very successful by this measure, these models do not seem configured to produce predictions about the magnitude of perceived harm. Rather, the theories tend to treat harm as either an antecedent to the "real action" (e.g., ref. 8) or as a fixed component within the model around which other factors can be organized (e.g., ref. 2). Given their focus on blame, this is entirely appropriate to these models; however, this state of affairs leaves open the possibility that unexpected factors may distort perceptions of harm.

Notably, recent work on harm magnification suggests that people may be prone to overestimate the damage caused by intentional harms. These effects apparently arise, at least in part, from blame motivation. However, like many explanations based on blame motivation, this argument has been based on indirect measures of the construct in question. Different bodies of work describe blame motivation in terms of the need to ascribe blame, to express moral condemnation, or to exact punishment (2, 4, 27, 28). In keeping with earlier research on harm magnification, we use the term "blame motivation" as convenient shorthand to refer to these needs collectively. This earlier research suggests that people seek to satisfy blame motivation when they encounter certain kinds of norm violations, and especially intentional harms (2, 3, 29). We test for direct evidence that intentional harm elicits this motivation. Such evidence, taken in conjunction with other findings, would constitute a more principled first step toward articulating the motivational component (if any) of these harm-magnification findings. More importantly, it would provide a set of concepts and tools that could be fruitfully applied to other research.

Previous Research in Harm Magnification

In a study that is representative of this prior work (30), participants read about a water shortage. Half the participants read that the shortage was caused intentionally and half read that it was caused unintentionally. Participants next saw a list of numbers (harms expressed in dollars, e.g., "medical supplies used: $80.00"). Their task was to estimate the sum of these numbers. Participants in the unintentional harm condition estimated accurately, whereas participants in the intentional harm condition overestimated the objective sum by a large margin, even if accuracy was incentivized, or if participants were told to simply give "the sum of the numbers you just saw" (see also refs. 6, 31, and 32 for related findings).

Other research (2–6, 15, 23, 26, 28–30) identifies blame motivation as a plausible psychological engine for this effect. However, it is difficult to say with certainty whether participants in these studies were truly motivated to blame, not least of all because the measures used to assay blame motivation were similar to those typically used in this area of inquiry: self-report measures of participants' desire to blame, punish, and condemn. Such reports may indeed reflect blame motivation; however, they could also be interpreted as judgments of blameworthiness rather than blame motivation. After all, people do find intentional harm-doing more blameworthy than unintentional harm-doing (e.g., refs. 27–29), and it is possible to think that an act warrants blame without actually experiencing motivation to ascribe blame.

As noted earlier, arbitrating between motivated accounts and epistemic accounts of judgment biases has been fraught with interpretational difficulty (23). Yet, early psychology studies using animal models routinely examined motivation, seemingly without encountering these problems. This work focused on behaviors related to choice and persistence (e.g., rats pressing levers to gain a food reward that they were motivated to obtain). The present investigation adapts the logic of this approach to moral judgment in humans. First, a preliminary experiment rules out an alternative explanation for previous findings, setting the stage for a more principled assay of motivational mechanisms. Exp. 1 allows participants to display sheer blame motivation, namely by choosing a task that involves blaming, condemning, and punishing over other appealing tasks, compared with a control condition. Exp. 2 also measures motivation, but with converging measures of persistence (effort, rate, and duration) in blaming.

Preliminary Experiment

Previous research [the "water shortage" experiment described above (30)] manipulated intentionality by comparing an intentional, person-caused harm with an unintentional, nature-caused harm. The comparison of intentional vs. unintentional causality was therefore confounded with the distinction between personal vs. natural causality. Examining situations of personal vs. natural causality was appropriate in these initial experiments, which were conducted with an eye toward addressing policy-related questions [resource allocation to natural disasters vs. person-caused disasters (33)]. A basic science perspective, however, should consider the personal/natural distinction as an alternative cause for the observed effect. Perhaps the effect is not the product of intentionality per se: Rather, people may overestimate human-caused harms, regardless of intent, perhaps because of a normative belief that human-caused harms do tend to be larger than otherwise-similar harms that arise by chance. After all, harm-doing humans can explicitly tailor their actions to be highly effective at producing harm, whereas natural disasters cannot.

We therefore sought to validate the previous findings with a completely different scenario that was free from this confound. If validated, this version of the task would provide a more principled foundation investigating blame motivation. A new scenario (described below) involved harm that was always person-caused.

If the "human vs. nature" account is correct, one should expect inflated harm estimates in both conditions. If however the "intentional vs. unintentional" account is correct, one should expect accurate estimates of unintentional harms, but exaggerated estimates of unintentional harms.

Following prior work (30), two groups of participants estimated the total magnitude of a set of harms (in dollars). The specific harms and associated dollar amounts were identical across conditions; however, these numbers were preceded by a vignette that framed these costs as accruing from either an intentional harm or an unintentional harm (randomly assigned). Specifically, the vignette described a nursing home employee giving residents inappropriate medications, either intentionally or unintentionally (the content of the vignette was based on various recent reports and media stories) (*Materials and Methods* and *Supporting Information*). After reading the vignette [identical reading times across conditions: $M_{(intentional)}$ = 53.9 s, $M_{(unintentional)}$ = 53.5 s], participants learned that they would see all of the resulting harms and how much they cost (in dollars), and that they would be asked about these harms and how much they cost (in dollars). Participants then saw these harms and associated dollar amounts and, immediately thereafter, estimated their sum.

As in previous experiments, estimates were accurate when the harm was framed as accidental [$4,557.20, SD = 2,261.70, vs. the correct answer of $4,433.43, $t(117)$ = 0.59, P = 0.55]. However, participants who viewed same set of numbers as arising from intentional harm-doing significantly overestimated their sum [M = $5,224.17, SD = 3,903.22, $t(130)$ = 2.32, P = 0.02, d = 0.20)] relative to ground truth. Adjusting for unequal variances, the two experimental conditions differed marginally from one another, $t(212.22)$ = 1.67, P = 0.097. It appears then that intentionality alone is sufficient to produce overestimates of objective harm.

In previous work, such results persist when studies incentivize accuracy, or when participants separately estimate punitive and compensatory damages (30). Science is cumulative however, and instead of seeking to repeat these demonstrations here, we proceeded to use this scenario to test the previously theorized motivational underpinnings of this effect. The fact that this scenario produced a weaker effect than previously observed could be a result of any of several factors, including the observation that psychological effects become smaller as the participant population becomes more experienced in psychological tasks (34, 35), or that participants were given more time to reflect on their responses than in previous experiments. In any case, if the present scenario constitutes a relatively weaker manipulation of the variable of interest, it provides a conservative platform for testing our main hypotheses.

Experiment 1. As noted, researchers—both advocates and critics of "motivated moralizing" theories—frequently assess blame motivation in terms of participants' propensity to assign degrees of blame, condemnation, and punishment. Applied as statistical mediators, these measures reveal covariance patterns that fit a model in which intentional harm engenders blame motivation, and blame motivation, in turn, exaggerates harm perception (30). However, this evidence does not say definitively whether these measures reflect actual motivation to blame, or simply participants' attempts to express their judgments of a harm-doer's blameworthiness. To address this problem, we sought a behavioral correlate of motivation. Animal researchers have often used free-choice behavior as a face-valid measure of their subjects' most basic motivations. For example, given the opportunity, healthy rats will often choose to expend effort gaining access to food over expending effort toward other ends. However, a rat that has been made cocaine-dependent, and is therefore motivated to obtain the drug, will often choose to expend effort to obtain cocaine, even when otherwise appetitive food is available. This change in free-choice behavior is an example

of a motivational shift in what the rat wants and how their new motivations shape their behavioral priorities [similar motivational shifts occur as a result of conditioning, surgery, genetic modification, and other manipulations (36–39)].

Applying this reasoning to human moral judgment (40), we suggest that if exposure to an intended harm induces the motivation to ascribe blame [as previously predicted by other researchers (2–6, 27–30)], people will freely choose to ascribe blame, perhaps even when an otherwise more appealing task is available. With this in mind, we constructed a simple test in which participants freely chose what task they would perform.

Participants read one of the two versions of the nursing home vignette shown to produce harm-magnification effects (randomly assigned). They then freely selected a second task to perform from among five options (*Materials and Methods*). Critically, one of these options was to assign blame, condemnation, and punishment to the harm-doer. If intentional harm elicits motivation to blame, condemn, and punish, then the tendency to blame, condemn, and punish should vary according to the intentionality of the harm.

Such sensitivity in fact emerged (Fig. 1). In the unintentional harm condition, the blame/condemn/punish task elicited only middling interest. However, in the intentional harm condition— that is, when the vignette was identical except that the medications were switched intentionally—selection of the "blame" task more than doubled, $\chi^2(1)$ = 6.7, P = 0.009. Although other tasks could have been similarly (or inversely) affected, the blame task proved to be uniquely sensitive to the intentionality manipulation: No other task differed across conditions. This pattern was corroborated in an overall χ^2 analysis of all intent/choice combinations, $\chi^2(4)$ = 11.4, P = 0.02. A third analytical approach of collapsing across the four control task options and using a 2(condition: intentional, unintentional) × 2(task choice: control, blame) χ^2 analysis yields the same result $\chi^2(1)$ = 8.2, P = 0.004.

Making sense of participants freely choosing one task over another demands saying something about what they wanted to do. The concept of "wanting" is inextricably linked to motivation. These data therefore support the notion that intentional harm-doing elicits some degree of motivation to blame, condemn, and punish the harm-doer.

Experiment 2. Motivation is sometimes expressed, not only in what people choose to do but also by the intensity or frequency with which they do it. Both people and animals who are motivated will tend to work longer and harder than those who are unmotivated (e.g., refs. 36 and 40). Following this logic, participants in Exp. 2 completed a reinforcement game modeled after classic studies of animal motivation (36–40). In a typical experiment of
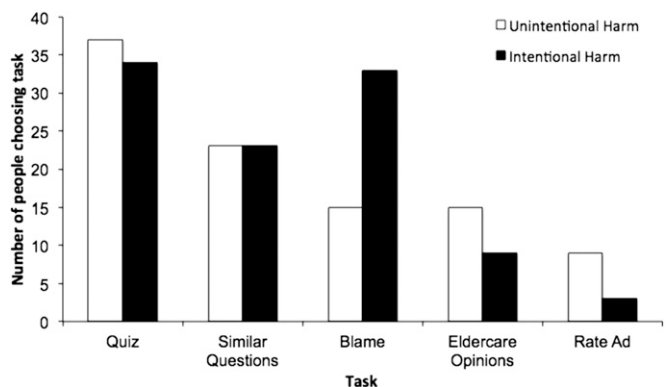

Fig. 1. Exp. 1: Number of participants selecting each task, separated by condition.

this type, the number of times that a rat pressed a lever indexed the rat's motivation to obtain a food reward associated with that lever (such that hungry rats would demonstrate greater effort and persistence in lever-pressing than nonhungry rats). The present experiment likewise indexed effort and persistence in humans—measuring mouse clicks rather than rat presses—to examine motivational responses to intentional and unintentional harms.

After reading the same vignette described previously (either the intentional or unintentional version, randomly assigned), participants proceeded to the reinforcement game. Participants could, if they wished, click on a blue envelope to "send a letter" to recruit signatures for a (pretend) petition. The petition called for the nursing home employee featured in the vignette to be investigated, and for appropriate actions to be taken. Participants could click as many or as few times as they wished (or not at all) to send increasing numbers of letters. Participants knew that the situation described in the vignette was a hypothetical composite, and that the petition was only part of the reinforcement game. However, games often do elicit genuine motivations, and these motivations often influence people's behavior within these contexts.

The instructions explained that not every letter sent would elicit a response. This feature of the design allowed the reward schedule to change over time, such that progressively more work (more clicking) was required to earn signatures as the task went on. Although "number of signatures earned" could be interpreted as a similar metric to a Likert-type scale (more signatures equating to higher blame ratings), this interpretation ignores the fact that participants had to do work to earn each signature. In keeping with classic work in behavioral reinforcement (36–39), we reasoned that doing more work indicated greater motivation.

Participants did substantially more work in the intentional harm condition ($M = 174.2$ clicks) than in the unintentional harm condition ($M = 25.6$ clicks), $t(67.59) = 6.0$, $P = 8.7 \times 10^{-8}$, corrected for unequal variances. This increase in effort was observed no matter how demanding the task became (that is, across all eight reward schedules examined). These results remained stable if counts were log-transformed [$M_{intentional} = 1.88$, $M_{unintentiona} = 0.72$, $t(114.28) = 8.1$, $P = 6.3 \times 10^{-13}$], or when time spent clicking, rather than number of clicks, was used as the unit of analysis [$M_{intentional} = 1:21$, $M_{unintentiona} = 0:51$, $t(84.12) = 3.4$, $P = .001$]. The intensity with which participants worked (click rate) was also four-times greater in the intentional harm condition [$M_{intentional} = 2.0$ clicks/second, $M_{unintentiona} = 0.5$ clicks/second, $t(84.85) = 7.4$, $P = 1.0 \times 10^{-10}$]. In sum, participants in the intentional harm condition worked more, faster, and for a longer period to assign blame, punishment, and moral condemnation.

We next addressed the question of how persistently participants worked as task demands increased. As noted, the task became more difficult over time, with an increasingly large number of clicks required to earn each signature. Motivated individuals should be expected to increase their level of effort to match the increasing task demands, whereas unmotivated individuals should be expected to decrease their level of effort to avoid the increasingly demanding task. Response elasticities (number of responses per unit price) (36) were computed to assess this relationship between effort and reward schedule (Fig. 2 and Fig. S1). A 2 (Condition: intentional, unintentional) × 8 (Reinforcement schedule: FR3–FR24) mixed ANOVA, with condition as a between-subjects factor and reinforcement schedules as a within-subjects factor, revealed a main effect of intent, with the intentional-harm condition eliciting more effort and persistence than the unintentional-harm condition, $F(1, 101) = 60.76$, $P = 4.5 \times 10^{-12}$, partial $\eta^2 = 0.374$. Consistent with previous research, a main effect of reinforcement schedule also emerged, confirming that response elasticities decreased as demandingness of the task increased (and, presumably, as the motivation to blame became more satisfied), $F(2.8, 40.2) = 63.96$, $P = 2.5 \times 10^{-30}$, partial $\eta^2 = 0.362$, Huynh–Feldt-corrected for nonsphericity. More
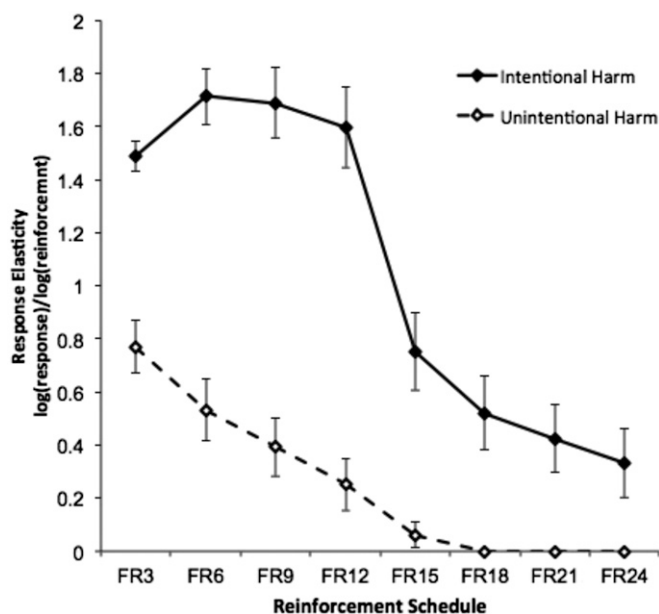

Fig. 2. Exp. 2: Motivation to earn signatures for mouse clicks (expressed as response elasticity) as a function of intentionality condition (intentional vs. unintentional harm). Horizontal axis shows fixed-ratio reinforcement schedule (for example, in FR3, every third response was reinforced). Error bars represent SEs.

importantly for present purposes, this relationship was qualified by a significant condition × reinforcement schedule interaction, $F(2.8, 300.89) = 15.89$, $P = 4 \times 10^{-9}$, partial $\eta^2 = 0.129$, Huynh–Feldt-corrected for nonsphericity, revealing that unmotivated participants consistently reduced their effort as task demands increased, whereas the motivated participants initially raised their level of effort to match the increasing task demands, then eventually decreased their effort (although not to the level of the unmotivated participants) as task demands became extreme and, presumably, as their motivation to blame became partly satisfied. This difference in the shape of the two groups' response functions is similar to the difference observed between drug-addicted animals vs. controls in drug self-administration (36). These findings align with the intuitive suggestion that when the going gets tough, the motivated keep going.

Discussion

These studies offer two general paradigms that address some outstanding questions concerning blame motivation in other research. Exp. 1 presents an example of how a general "choose your task" paradigm can assay motivation in a face-valid way. Exp. 2 provides a template for measuring motivated judgment according to classic experimental criteria for measuring motivation (effort, rate, duration). A mouse-clicking paradigm has the benefit of being easily deployable online, and we include our code to help researchers capitalize on this (Dataset S1). Although online data collection allows for larger and more diverse samples than what most psychological scientists can practically acquire in the laboratory, a persistent criticism of online experimentation has been its apparent inability to capture actual behavior. However, the paradigm introduced here captures precisely the same sort of behavior that has been used to index motivation in laboratory settings for decades. The simplicity of lever-presses or button-presses is part of the appeal of these measures. Of course, these measures would be best applied in conjunction with in-laboratory experiments that test whether similar results can be obtained using ecologically meaningful responses.

With respect to the harm magnification effect, the present studies help to substantiate prior claims that people are motivated to blame intentional harm-doers (29, 30, 41). These studies go beyond prior meditational evidence, to experiments measuring motivation with converging measures of choice and persistence (including effort, rate, and duration). Future work in this area should further clarify the proposed causal relationship between blame motivation and harm magnification. One challenge to this research is that the actions of choosing to blame and persisting in blaming may themselves satisfy the very motivations in question. At present, results are consistent with the suggestion that intentional harm-doers incite blame motivation, and blame motivation causes people to overestimate the damage done. If this model is correct, it would align with existing research in other domains. This research reveals that other kinds of magnitude biases arise from motivation (19), and that the perceived intentionality of agents can influence people's predictions about numerical magnitude in probabilistic settings (42).

Perceived intent emerges as catalyzing a motivated social cognitive process related to social prediction and control. Intent matters crucially to social cognition: intent for good or ill is one of the earliest and most basic judgments people make about each other. If intentional harms seem worse than they truly are, society may spend more money on them than on objectively more damaging harms (33, 43). A possible motivated account of this phenomenon acknowledges the nonrational biases in damage estimates and potential impact on policy priorities.

To be sure, all vignette-based studies are limited by the particulars of vignette. For the specific purpose of understanding harm-magnification effects, future work will need to construct similar test cases using a broader range of scenarios. Previous work has found harm magnification effects across various stimulus materials and harm types, including videos, short-story excerpts, adaptations of classic psychology tasks, objective financial harms, emotional harms, and "harm" as defined idiosyncratically by participants, and in neurotypical and patient populations (29, 30, 41). However, our purpose here was primarily one of introducing a new approach for studying a topic that has recently captured broad interest but for which no adequate test seemed to be available. Applying the concepts of choice and persistence to motivation more broadly may help to illuminate answers that have, to this point, remained obscured by the interpretational ambiguity of other measures.

Materials and Methods

Participants and Recruitment. Across the three studies, 630 participants were recruited using Amazon Mechanical Turk and then redirected to an experimental website [preliminary experiment: $n = 309$, $M_{age} = 30$ y (10.5), 121 male, 187 female, 1 not reported; Exp. 1: $n = 201$, $M_{age} = 34$ y (13.5), 1 not reported, 90 male, 111 female; Exp. 2: $n = 120$, $M_{age} = 33$ y (12.2), 51 male, 68 female, 1 prefer not to say]. Participants were paid standard rates in exchange for their participation (preliminary experiment: $0.40, $6.26/h; Exp. 1, $0.40, $5.39/h; Exp. 2: $0.40, $5.04/h; pay rate for Exp. 2 is calculated from Amazon Mechanical Turk acceptance/submission time rather than actual task start/end time, and is therefore likely an underestimate). Recruitment was limited to the United States. Recent research suggests that samples recruited via Mechanical Turk are more demographically and cognitively representative of national distributions than are traditional student samples, with data reliability typically being at least as good as that obtained via traditional sampling (44–46). Participants who had completed any previous study in this line of work were automatically prevented from accessing the recruitment page (47). Across all three studies, 7.5% of participants were excluded for failing the manipulation check or basic checks for meaningful responses (full details in *Supporting Information*).

Vignette. The framing vignette described a nursing home employee who mixed up patients' medications (see *Supporting Information* for full text). This vignette was a composite based on various academic papers, government reports, and news releases concerning high rates of inappropriate medication in nursing homes (e.g., nursing home employees giving

antipsychotics without any clinical indication for this treatment, or ignoring maximum recommended daily doses of these medications) (48–50). From these reports, which included instances of both intentional and unintentional distribution of inappropriate medication, we constructed a composite narrative, in which inappropriate medication was dispensed either intentionally or unintentionally (50/50 random assignment). Across conditions, the vignette featured the same act (mixing up the pills), committed by the same person (an employee named Jake), resulting in the same harms (symptoms and associated medical costs), to the same people (nursing home residents). The proximal cause of the to-be-estimated harms was also the same across conditions (inappropriate medication). Only Jake's intent varied between conditions. Participants were fully informed that this vignette was a composite, rather than a veridical report of a specific instance: "The following story is a composite of several true events. We are not asking about any specific instance. Resemblance to any specific person(s) or location(s) is coincidental."

Preliminary Experiment: Harm Magnification. Data were analyzed using Microsoft Excel 2011 for Mac and PASW Statistics 18. Of 310 participants recruited, after reading the consent, 309 volunteered. Of these, 27 failed to complete the task. Of the remaining 282 participants, 9.6% were excluded for failing the manipulation check or basic checks for meaningful responses. For complete information on these checks and the number of persons excluded by each one, see Participant Exclusion Criteria in *Supporting Information*.

After reading the composite vignette (full text in *Supporting Information*), participants were informed that they would next see all of the harms arising from the actions describing the vignette, along with the associated medical costs (in dollars). They were further told that this information would go by quickly, and that they should pay close attention, as they would be asked to report these costs. Participants then viewed eight consequences of the medication mix-up, along with their associated costs (e.g., "overtime staff hours: $271.10"). The items were presented for 2,000 ms each, and totaled $4,433.43. After all eight items had been presented, participants were asked to "Estimate the TOTAL cost (i.e., the sum of the numbers you just saw)." Participants were given 15 s to make this estimate (more than the 10 or 12 s allowed in previous studies, allowing more time for reflective processing). After making their estimates, participants completed a manipulation check as to the intentionality of the harm: "Did Jake give the wrong medications on purpose, or on accident?" A reading check asked "Did all of the residents recover in the end?" Participants then had the option of giving their opinions as to whether Jake should lose his job (60% "yes," 38% "no," 2% "not sure/no opinion"), and whether it would be better or worse for the residents to know what caused the uptick in medical issues, assuming that no such mix-up ever occurred again (51% "they'd be better off knowing," 40% "they'd be better off not knowing," 9% "not sure/opinion"). Finally, participants reported their demographic information.

Experiment 1: Choice. With 200 participants recruited, 201 participants completed the task, possibly because a 201st participant began the task before the 200th participant finished the task, allowing the extra participant to join the study before being "locked out." All 201 participants were retained. Data were analyzed using Microsoft Excel 2011 for Mac, PASW Statistics 18, and Qutantsci software (41).

After reading the vignette, participants answered three questions before proceeding to the reinforcement game: "Did Jake give the wrong medications on purpose, or on accident?" (manipulation check); "Do you feel like you remember hearing/reading about an event like this (e.g., on the news)?" (5% "yes," 26% "maybe," 69% "no"); and "Assume that Jake never mixed up pills again, and that there was no long-term harm done. Do you think that the residents of the nursing home would be better off knowing why they got sick, or not knowing why they got sick?" (62% "they'd be better off knowing," 28% "they'd be better off not knowing," 10% "not sure/opinion"). Seven participants (3%) answered the third question (manipulation check) incorrectly. Their inclusion or exclusion was inconsequential, and their data were retained in the analysis reported above [if excluded, effect of intent on selection of "blame/condemn/punish" task: $\chi^2(1) = 6.8$, $P = 0.009$, all other conditions not significant; overall 2×5 χ^2: $\chi^2(4) = 11.5$, $P = 0.02$].

Next, participants completed the main task, and then reported their demographic information (age, sex, and race). In the main task, participants saw a page with the following instructions: "Thanks—you're about 75% done. Please choose one of the following to complete (pending availability). They all take about the same amount of time." Participants chose freely among the following five options, presented in a random order:

Offer your opinion about how (if at all) someone like Jake should be punished for what he did, and what amount of blame/moral condemnation he deserves.

Take a short quiz about the cost of healthcare in the U.S. (answers will be provided at the end of the quiz).

View a healthcare advertisement and offer your opinions about its tactics and effectiveness.

Offer your opinion about elder care in the U.S.

Answer various questions about nursing home situations like one you just read about.

Participants selected one task to complete. Propensity to select the "blame/punish/condemn" option (presented first above, but again, presented in a new random order to each participant) was the critical dependent variable. The other four options controlled for various alternative explanations. For example, participants' tendency to select the "blame/punish/condemn" option more in the "intentional harm" condition than in the "unintentional harm" condition could, in principle, arise because the "intentional harm" condition simply made participants more concerned about eldercare, making people to want to express their feelings about this, but without actually generating any motivation to blame. However, this interpretation would predict an increase in selections of both the "blame/punish/condemn" option and the "opinions about elder care" option, a pattern that was not observed. Another possibility might be that reading a vignette about a healthcare-related situation could make the concept of healthcare accessible, thereby increasing processing fluency for tasks related to healthcare, thus increasing participants' tendency to select the predicted option; however, because all five options relate to healthcare, it is unclear how this explanation would predict the unique increase in participants' desire to blame/condemn/punish. In any case, the uniqueness of the blame task's sensitivity to intent was borne out statistically in the analyses above, regardless of whether a more liberal, uncorrected alpha ($\alpha = 0.05$) or a more conservative, Bonferroni-corrected α ($\alpha = 0.01$) was used to evaluate participants' preference for each of the five activities.

Experiment 2: Effort and Persistence. Although 120 participants were recruited, 121 data rows were created. However, one row apparently recorded no analyzable data. The cause of the error is unknown. Three participants were excluded as statistical outliers for clicking a number of times greater than 3 SD above the condition mean (the highest excluded figure being 857 clicks). Results were identical if these three participants were retained.

After reading the vignette, participants answered two questions before proceeding to the reinforcement game: "Do you feel like you remember hearing/reading about an event like this (e.g., on the news)?" (4% "yes," 7% "maybe," and 89% "no") and "Did Jake give the wrong medications on purpose, or on accident?" (manipulation check). Four participants (3%) answered the manipulation check incorrectly. Their inclusion or exclusion was negligible, and their data were retained in the analyses reported above.

After answering the familiarity and manipulation check questions, participants proceeded to the main task. At the top of the screen was written "Your opinion: Should Jake be investigated? Punished?" Also onscreen were task instructions, a picture of a petition (with twenty blank spaces for signatures), and a picture of a blue envelope. Above the envelope was written "Click envelope repeatedly to try to get signatures." The task instructions read as follows:

If you think that Jake should receive some blame, punishment, and/or moral responsibility for what happened, please indicate this by clicking on the blue envelope below. Each time you click on the envelope, you will "send a letter" to recruit a signature for the (pretend) petition, signaling that Jake should be investigated and appropriate actions should be taken.

Just like in real life, not every letter will get a response. The more signatures you get, the more strongly we assume you feel that Jake should be investigated, blamed, punished and/or held morally responsible.

The game began with a fixed-ratio reinforcement schedule of 3 (FR3), in which one out of every three clicks produced a signature. Which of the three clicks in each triad triggered the signature was randomized (i.e., in clicks 1–3, the third click might produce the signature; in clicks 4–6, the fifth might produce the signature). Every time participants earned five signatures, the reinforcement ratio increased by three, progressing through FR6, FR9, FR12, FR15, FR18, FR 21, and FR24. Participants were free to end the game at any time, and knew that they would earn no more or less money as a function of how long they spent clicking (see Fig. S2 for p. 2 of a filled petition).

After the main task, participants were presented with two optional questions: whether they had any opinion about what should happen to the person who mixed up the medication, and whether they thought it would be better or worse for the residents to know what caused the sudden uptick in medical issues, assuming that the incident was never repeated. Finally, demographic information was collected.

The analyses of time spent clicking and click rate include only participants who clicked at least once (0 clicks is valid data for the count-based analysis, but 0 clicks in 30 s is uninformative for click rate, and including such data would have biased results in a direction consistent with our hypotheses). Time spent clicking and click rate were indexed by the times at which participants began and ended the task. This approach underestimates true click rate, because participants presumably spent part of this time reading instructions and orienting to the task. In principle, it is possible that participants in one condition took longer to begin clicking, or remained in the task for longer after they were done clicking. However, this cannot explain the results observed in this experiment. If participants in the intentional harm condition spent longer on the task page without actually clicking more, this could indeed produce an artificial "time spent clicking" difference; however, this explanation is inconsistent with the observed differences in "number of clicks" and "click rate" (it would predict the opposite of the observed results). Likewise, if participants in the unintentional harm condition spent more time on the task without actually clicking any less, their click rate would be artificially deflated in our analysis; however, this is inconsistent with the observed difference in number of clicks, and would suggest that the true effect of condition on time spent working is, in fact, even larger than what we report.

1. Alicke MD (2008) Blaming badly. *J Cogn Cult* 8(1-2):179–186.
2. Alicke MD (2000) Culpable control and the psychology of blame. *Psychol Bull* 126(4):556–574.
3. Ditto PH, Pizarro DA, Tannenbaum D (2009) Motivated moral reasoning. *Psychol Learn Motiv* 50:307–338.
4. Haidt J (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol Rev* 108(4):814–834.
5. Nadler J, McDonnell MH (2012) Moral character, motive, and the psychology of blame. *Cornell Law Rev* 97(2):255–304.
6. Sood AM, Darley JM (2012) The plasticity of harm in the service of criminalization goals. *Calif Law Rev* 100(5):1313–1358.
7. Guglielmo S, Monroe AE, Malle BF (2009) At the heart of morality lies folk psychology. *Psychol Inq* 52(5):449–466.
8. Malle BF, Guglielmo S, Monroe AE (2014) A theory of blame. *Psychol Inq* 25:147–186.
9. Greene JD (2009) Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Exp Soc Psychol* 45(3):581–584.
10. Greene JD (2014) *Moral Tribes: Emotion, Reason and the Gap Between Us and Them* (Penguin, New York, NY).
11. Pettit D, Knobe J (2009) The pervasive impact of moral judgment. *Mind Lang* 24(5):586–604.
12. Kelley HH (1971) *Attribution in Social Interaction* (General Learning Press, New York, NY).
13. Pizarro DA, Tannenbaum D (2011) Bringing character back: How the motivation to evaluate character influences judgments of moral blame. *The Social Psychology of Morality: Exploring the Causes of Good and Evil*, eds Mikulincer M, Shaver PR (APA Press, Washington, DC), pp 91–108.
14. Tannenbaum D, Uhlmann EL, Diermeier D (2011) Moral signals, public outrage, and immaterial harms. *J Exp Soc Psychol* 47(6):1249–1254.
15. Lerner MJ (1965) Evaluation of performance as a function of performer's reward and attractiveness. *J Pers Soc Psychol* 1(4):355–360.
16. Lerner MJ (2003) The justice motive: Where social psychologists found it, how they lost it, and why they may not find it again. *Pers Soc Psychol Rev* 7(4):389–399.
17. Gray K, Young L, Waytz A (2012) Mind perception is the essence of morality. *Psychol Inq* 23(2):101–124.
18. Gray K, Wegner DM (2009) Moral typecasting: Divergent perceptions of moral agents and moral patients. *J Pers Soc Psychol* 96(3):505–520.
19. Schein C, Gray K (2014) The prototype model of blame: Freeing moral cognition from linearity and little boxes. *Psychol Inq* 25(2):236–240.
20. Balcetis E, Dunning D (2006) See what you want to see: Motivational influences on visual perception. *J Pers Soc Psychol* 91(4):612–625.

21. Custers R, Aarts H (2010) The unconscious will: How the pursuit of goals operates outside of conscious awareness. *Science* 329(5987):47–50.
22. Bargh JA (2002) Losing consciousness: Automatic influences on consumer judgment, behavior, and motivation. *J Consum Res* 29(2):280–285.
23. Kahneman D (2011) *Thinking, Fast and Slow* (Macmillan, Chicago, IL).
24. Kunda Z (1990) The case for motivated reasoning. *Psychol Bull* 108(3):480–498.
25. Ross M, Fletcher GJO (1985) Attribution and social perception. *The Handbook of Social Psychology* (3rd ed). eds Lindzey G, Aronson E (Addison-Wesley, Reading, MA) pp 73–122.
26. Tetlock PE, Levi A (1982) Attribution bias: On the inconclusiveness of the cognition-motivation debate. *J Exp Soc Psychol* 18(1):68–88.
27. Goldberg JH, Lerner JS, Tetlock PE (1999) Rage and reason: The psychology of the intuitive prosecutor. *Eur J Soc Psychol* 29(56):781–795.
28. Robinson PH, Darley JM (1995) *Justice, Liability, and Blame: Community Views and the Criminal Law* (Westview Press, Boulder, CO).
29. Darley JM, Pittman TS (2003) The psychology of compensatory and retributive justice. *Pers Soc Psychol Rev* 7(4):324–336.
30. Ames DL, Fiske ST (2013) Intentional harms are worse, even when they're not. *Psychol Sci* 24(9):1755–1762.
31. Pizarro DA, Laney C, Morris EK, Loftus EF (2006) Ripple effects in memory: Judgments of moral blame can distort memory for events. *Mem Cognit* 34(3):550–555.
32. Ward AF, Olsen AS, Wegner DM (2013) The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychol Sci* 24(8):1437–1445.
33. Gilbert D (2011) Buried by bad decisions. *Nature* 474(7351):275–277.
34. Rand DG, et al. (2014) Social heuristics shape intuitive cooperation. *Nat Commun* 5:3677.
35. Chandler J, Mueller P, Paolacci G (2014) Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behav Res Methods* 46(1):112–130.
36. DeGrandpre RJ, Bickel WK, Hughes JR, Layng MP, Badger G (1993) Unit price as a useful metric in analyzing effects of reinforcer magnitude. *J Exp Anal Behav* 60(3):641–666.
37. Ferster CB, Skinner BF (1957) *Schedules of Reinforcement* (Appleton-Century-Crofts, New York, NY).
38. Lattal KA, Gleeson S (1990) Response acquisition with delayed reinforcement. *J Exp Psychol Anim Behav Process* 16(1):27–39.
39. Hursh SR, Raslear TG, Shurtleff D, Bauman R, Simmons L (1988) A cost-benefit analysis of demand for food. *J Exp Anal Behav* 50(3):419–440.
40. Morewedge CK, Huh YE, Vosgerau J (2010) Thought for food: Imagined consumption reduces actual consumption. *Science* 330(6010):1530–1533.
41. Freeman D, Evans N, Černis E, Lister R, Dunn G (2015) The effect of paranoia on the judging of harmful events. *Cogn Neuropsychiatry* 20(2):122–127.
42. Caruso EM, Waytz A, Epley N (2010) The intentional mind and the hot hand: Perceiving intentions makes streaks seem likely to continue. *Cognition* 116(1):149–153.
43. McGraw AP, Todorov A, Kunreuther H (2011) A policy maker's dilemma: Preventing terrorism or preventing blame. *Organ Behav Hum Dec* 115(1):25–34.
44. Buhrmester M, Kwang T, Gosling SD (2011) Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspect Psychol Sci* 6(1):3–5.
45. Mason W, Suri S (2012) Conducting behavioral research on Amazon's Mechanical Turk. *Behav Res Methods* 44(1):1–23.
46. Paolacci G, Chandler JJ, Ipierotis P (2010) Running experiments on Amazon Mechanical Turk. *Judgm Decis Mak* 5(5):411–419.
47. Peer E, Paolacci G, Chandler JJ, Mueller PA (2012) Selectively recruiting participants from Amazon Mechanical Turk using Qualtrics. Available at SSRN: ssrn.com/abstract=2100631. Accessed March 13, 2013.
48. Centers for Medicare and Medicaid services (2013) *Press Release: New Data Show Antipsychotic Drug Use Is Down in Nursing Homes Nationwide*. Available at www.cms.gov/newsroom/mediareleasedatabase/press-releases/2013-press-releases-items/2013-08-27.html. Accessed February 9, 2014.
49. Chen Y, et al. (2010) Unexplained variation across US nursing homes in antipsychotic prescribing rates. *Arch Intern Med* 170(1):89–95.
50. Span P (2010) Overmedication in the nursing home. *The New York Times*. Available at newoldage.blogs.nytimes.com/2010/01/11/study-nursing-home-residents-overmedicated-undertreated/. Accessed February 9, 2014.