

Responsibility Amplifies Empathic Forecasts

Fausto J. Gonzalez and Minah H. Jung
New York University

Clayton R. Critcher
University of California, Berkeley

Inspired by theoretical and empirical work on emotion, psychological distance, moral psychology, and people's tendency to overgeneralize ecologically valid relationships, 3 studies explore whether, why, and for whom responsibility amplifies empathic forecasts (RAEF)—the perception that an intentional agent's social actions will produce stronger affective responses in others than if those same outcomes were to occur randomly or unintentionally. In Study 1, participants thought that pleasant or aversive videos would elicit stronger reactions when participants themselves (instead of the random determination of a computer) would select the video another would watch. This was explained by responsible agents' own stronger reactions to the stimuli. Study 2 identified what about agents' responsibility amplifies empathic forecasts: the combination of clearly causing and intending the other's outcome. Study 3 demonstrated that RAEF need not extend to all responsible agents equally. Participants considered how to divide (vs. how another participant would divide or how a computer would randomly split) \$10 with a recipient. In this context, we found the weight of causal responsibility looms larger in the self's mind when the self is responsible for the recipient's fate than when another responsible agent is. Furthermore, the self thought that the recipient's emotional reaction would be more strongly influenced by the size of the self's own (compared to another's or a computer's) allocation decision. The Discussion focuses on how RAEF relates to other models connecting agency and experience, provides initial evidence that RAEF need not be egocentric, and identifies open questions that remain for future research.

Keywords: social cognition, empathic forecasting, causality, intent, egocentrism

Supplemental materials: <http://dx.doi.org/10.1037/xge0000976.supp>


When contemplating decisions, people often consider how others will react. How much delight will a gift bring a friend? How much annoyance will one's tardiness cause in a coworker? In each case, one person has an effect on another's emotional state. We explore a feature that affects such forecasts: whether the agent is clearly responsible for the other's fate.

Several decades of research on affective forecasting has focused on people's (in)ability to forecast their own future hedonic states

(Baron, 1992; Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; Loewenstein & Schkade, 1999; Kahneman & Snell, 1990; van Dijk, Van Dillen, Seip, & Rotteveel, 2012; Wilson & Gilbert, 2005; Wilson, Wheatley, Meyers, Gilbert, & Axsom, 2000). More recently, researchers have examined not only how forecasters predict their own emotional states (affective forecasting), but also how they predict the emotional reactions of others (empathic forecasting). One general lesson is that many of the same biases seen in affective forecasts extend to empathic forecasts. For example, people overestimate—both for themselves and for someone else—how upsetting it would be to fail a test (Pollmann & Finkenauer, 2009) or how sad it would be to learn one's partner had lied (Green et al., 2013). Affective forecasts often correlate with (and thus may inform) empathic forecasts, whether for close others or even strangers (Pollmann & Finkenauer, 2009). After all, the self serves as a starting point or anchor for social forecasts (Critcher, Dunning, & Rom, 2015; Lau, Morewedge, & Cikara, 2016; Van Boven & Loewenstein, 2003; see also Gilovich, Savitsky, & Medvec, 1998).

Although empathic and affective forecasts frequently converge, in other cases they diverge. For example, the self may be differentially aware of its own as opposed to others' coping resources. This explains why even though people think they would be just as upset to fail a driver's exam as would another, they also expect that four days later they themselves would be emotionally coping much better than another (Igou, 2008). But

This article was published Online First October 29, 2020.

 Fausto J. Gonzalez and Minah H. Jung, Department of Marketing, Leonard N. Stern School of Business, New York University; Clayton R. Critcher, Haas School of Business, University of California, Berkeley.

The research reported in this article was supported in part by U.S. National Science Foundation Award 1749608 to Clayton R. Critcher. We thank Taylor Loskot, Natalie Abber, Jessica Compartore, Alondra Tlatelpa, Nora Harhen, David Lee, Joshua Woznica, and Rowena Zhang for their assistance with data collection.

All code, data, and materials are publicly available through the Open Science Framework (<https://osf.io/apbvs/>). This research was previously presented at the Society for Consumer Psychology's annual meeting in San Francisco, California in March 2017.

Correspondence concerning this article should be addressed to Fausto J. Gonzalez, Department of Marketing, Leonard N. Stern School of Business, New York University, 40 West 4th Street, Tisch 915, New York, NY 10012. E-mail: fjg4@stern.nyu.edu

in contemplating extreme outcomes for which people do not know how they themselves would adapt—like suffering a major disability—people not only overestimate how negatively they would be affected (Ubel, Loewenstein, Schwarz, & Smith, 2005), but they assume they would fare even worse than others (Walsh & Ayton, 2009).

In other cases, empathic forecasters struggle to make good use of social category cues about their targets. For example, forecasters exaggerated the emotional impact that winning or losing U.S. midterm elections would have on Republicans and Democrats, respectively. Empathic forecasters fared much better when they merely forecasted the reactions of citizens on the winning and losing sides, with their political affiliations unspecified (Lau et al., 2016). In other research, White and Black participants alike thought that White targets would feel more social pain in response to slights (e.g., being made fun of by a friend) than would matched Black targets, even though White and Black participants reported these experiences to be equally painful (Deska, Kunstman, Bernstein, Ogungbadero, & Hugenberg, 2020; see also Hoffman, Trawalter, Axt, & Oliver, 2016). Finally, forecasters often lean on stereotypes (e.g., that an upcoming Black interaction partner would be especially likely to become angry), which can influence these forecasters' own behavioral intentions even before any actual interaction begins (Moons, Chen, & Mackie, 2017). In every case, social category information either led empathic forecasters astray or encouraged them to approach the other with different and even counterproductive inclinations (e.g., to withhold social support; Deska et al., 2020).

In considering the state of the empathic forecasting literature, Green et al. (2013) identified an omission: Little attention had been paid to how the self's own actions would affect others emotionally. The researchers examined emotional reactions to transgressions in romantic relationships. In such contexts, one can differentiate the perpetrator from the victim: the one who commits the transgression or is transgressed against, respectively. In a longitudinal design, Green et al. (2013) replicated a robust finding in the affective and empathic forecasting literatures: Perpetrators and victims reported that actual transgressions made them less sad than they had forecasted. Furthermore, although forecasters were correct that they themselves would be sadder as a perpetrator than a victim, they mistakenly thought their partner would show the reverse pattern.

These results are intriguing in part because they hint at a phenomenon that this article more directly and generally explores. Perpetrators—those responsible for victims' states—assumed that their actions would have a particularly strong effect on that other. Furthermore, in imagining being responsible for hurting another, perpetrators anticipated being similarly strongly emotionally moved. Green et al. (2013) refer to this co-occurrence as event projection: Perpetrators use their own (anticipated) elevated affective response to inform the anticipated affective experience of those they are affecting.

In this article, we ask whether responsibility—causing an outcome through intentional actions—influences empathic forecasts. Although Green et al.'s (2013) data provide initial support for the plausibility of our ideas, note that they do not offer a direct test. That is, in their paradigm, there is no comparison by which victims experience an equivalent outcome that is not visited upon them by a causal, intentional agent. (Of course, transgressions within ro-

mantic relationships do not offer an ideal context to test this hypothesis, nor was this the goal of Green and colleagues' research). In what follows, we develop why we think responsibility—even when targets are blind to it—amplifies empathic forecasts.

Empirical Precedent for a Connection Between Responsibility and Emotional Impact

We draw on three research traditions that converge in supporting the plausibility of our hypothesis that responsibility will amplify empathic forecasts. First, we argue that responsibility draws agents closer to those they affect, and such proximity is known to amplify the emotional evocativeness of actions. Second, we reason by analogy from the moral psychology literature that intentional, causal agents are believed to be emotionally impactful. Third, we highlight that people often do have amplified responses when they are acted upon by responsible agents, a relationship that may be overgeneralized (as ecologically valid connections often are) to cases in which such responsibility is hidden from targets.

Responsibility Draws Agents Closer to Outcomes

Stimuli often lose their emotional punch with greater distance. As emotion regulation researchers have shown, such distance can take various forms. Reconstructing evocative episodes from a third-person perspective (Kross & Ayduk, 2008; Leitner et al., 2017), pretending an arousing object is hypothetical (McRae, Ciesielski, & Gross, 2012), or literally creating physical distance (Gable, Reis, & Elliot, 2000) can all blunt the emotional evocativeness of stimuli. Even when one does not literally move away from a stimulus, but merely imagines doing so, stimuli elicit less emotion (Davis, Gross, & Ochsner, 2011).

Responsibility creates subjective proximity between agents and the outcomes they produce. Being personally responsible for an event gives rise to a sense of agency with a distinct phenomenological signature (Haggard & Tsakiris, 2009). Crucial to our argument, actively bringing about outcomes actually leads people to encode the cause–effect sequence as a temporally compressed unit (Haggard, Clark, & Kalogeras, 2002; Pfister, Obhi, Rieger, & Wenke, 2014). In other words, agents feel even closer than is objectively warranted by those events they try to bring about. This means the emotional significance of an agent's actions may become more vivid and intense.

Although the self's proximity to an emotionally laden event may amplify the self's own response to it (Greene, Nystrom, Engell, Darley, & Cohen, 2004), there is also evidence that the proximity of agents (more generally) to those they affect can amplify such actions' emotional evocativeness. As one example, when an agent harms another through direct physical contact—thereby creating as much proximity between the agent and patient as possible—judgments of moral impermissibility increase (Cushman, Young, & Hauser, 2006). And given perceptions of moral impermissibility are generally affect-backed (Bartels, 2008; Cushman, Gray, Gaffey, & Mendes, 2012; Hume, 1740; Nichols, 2002), it is a short leap to assume such agent–victim proximity encourages affective discomfort in the self as well. More generally, the properties that constrain the self's own decision making often apply to the self's encoding of others'

decisions as well (Cushman, 2015). In combination, this suggests that responsibility—given it draws agents closer to those they affect—may enhance the emotional evocativeness of such outcomes, thereby elevating empathic forecasts.

Moral Psychology Has Linked Responsible Agents With Experiential Patients

More generally, we suggest that people possess a basic social-cognitive template that links causal agency and emotional reactivity that assists with understanding their social worlds. The sharpest theoretical articulation of related ideas—as well as the clearest empirical support for them—comes from moral psychology. In particular, this literature suggests mutually reinforcing relationships among responsibility (in particular, causality and intentionality), blame, and perceived emotional impact. We review this perspective, which has been used to explain how people understand moral transgressions, in order to illustrate how it may inform the role of responsibility in empathic forecasting more generally.

According to Gray, Waytz, and Young (2012), people—even those with differences in their moral values (Schein & Gray, 2015)—understand immoral actions using the same cognitive template: An agent intentionally causes harm to a vulnerable and emotionally experiential patient (Schein & Gray, 2018). Most straightforwardly, this template is used to recognize what actions are morally wrong. But much as the mind automatically fills in an occluded portion of a visual field (based on top-down knowledge of what is being seen), the template also guides inferences about what has fully transpired when immorality is witnessed. This process is called dyadic completion (Gray, 2012) and is one of a number of reasons why moral perception has been identified as analogous to visual perception (Schein, Hester, & Gray, 2016).

Most relevant to the present purposes, this dyadic template encourages perceivers to look for and even see emotional suffering merely upon learning a causal agent commits an act one thinks is wrong (Gray & Schein, 2012). Schein and Gray (2018) call this *patient dyadic completion* and identify it as the most important way that the template is used to fill in or complete social perceptions. It is why norm violations compel people to find suffering victims, even when such a sufferer may be hard to locate (DeScioli, Gilbert, & Kurzban, 2012; Gray, Young, & Waytz, 2012; Schein, Goranson, & Gray, 2015). After all, if causal agents indeed do wrong by creating suffering in victims, then knowing a causal agent has done wrong is tantamount to knowing she has caused suffering. Although we certainly concur that this link among causality, intentionality, and emotional experience applies to lay understandings of immorality (see also Alicke, 2000; Malle, 2006), we also suggest that this may reflect a more general template by which people understand the link between responsibility and emotional consequence.

Because actions that are morally wrong are assumed to produce even greater emotional consequences, it is useful to understand what features of actions increase moral condemnation. And given our goal is to move beyond an examination of moral transgressions to emotionally impactful social actions more generally, it will be particularly useful to identify nonmoral features of actions that fit this bill. Fortunately, Cushman and Young (2011) have already done much of this heavy lifting by identifying two such components that, in tandem, compose responsibility. One component is

intent. Judgments of intentions are what bridge certain features of moral actions (e.g., causing a harm as a means to an end as opposed to a side effect of another action; Cushman & Young, 2011) and a desire to blame (Ames & Fiske, 2015). Coming full circle to our own interest in whether responsibility amplifies empathic forecasts, such desires may lead people to claim that actors actually did more harm and to misremember that actors exacted more damage than the objective evidence reflected (Ames & Fiske, 2013). The second component is clear evidence of causality. For example, when one brings about a harm through a direct action (a commission) instead of a failure to act (an omission), one is seen to be more clearly the cause of that action. As a result, more blame is offered (Cushman & Young, 2011). According to dyadic completion, such elevated blame is itself a cue to elevated emotional impact. These features—causality and intentionality—define immorality most clearly when they operate in tandem (see Greene et al., 2009). We reason that these features—especially when combined—may cue the emotional impactfulness not only of immoral behaviors but social actions more generally.

Overgeneralization of an Ecologically Valid Relationship

Finally, we note that intentional, causal agents often do have elevated emotional impacts. Most basically, we know that the self's responsibility for its own outcomes can induce elevated emotional responses in the self. Achievements seen to result from one's own efforts are more likely to cause pride than those for which the self does not feel responsibility (Weiner, 1985). Furthermore, acts of commission—those for which perceived causality and thus responsibility is heightened (Spranca, Minsk, & Baron, 1991)—prompt more regret (at least in the short term) than do acts of omission (Gilovich & Medvec, 1995).

In addition, one's responsibility for others' outcomes can elevate others' emotional reactions. When participants thought they missed out on a financial reward because their partner intentionally put in low effort, they were more unhappy than if the reason for failure was ambiguous (Arditte Hall, Joormann, Siemer, & Timpano, 2018). Gray and Wegner (2008) found that those who received otherwise equivalent electric shocks reported them to be more painful when they were supposedly administered by a human than by a computer. The very meaning of experience changes when it is the result of an intended social action. Learning that the freshly baked cookies left on one's doorstep were actually intended for one's neighbors turns a heartwarming gesture into an uncomfortable dilemma.

In these social examples, the recipients know (or learn) that their fate was or was not intentionally caused by a fellow human. But what if such knowledge were hidden from recipients? Much as dyadic completion argues that people see suffering when none is objectively present, we suggest that empathic forecasters may similarly overgeneralize. They may assume stronger emotional reactions even when the reason for such elevated responses (i.e., responsibility) is hidden from targets. After all, both the heuristics and biases as well as the social judgment literatures are filled with examples of people overgeneralizing ecologically valid relationships to contexts to which they should not apply (Gigerenzer & Brighton, 2009; Reit & Critcher, 2020; Steinmetz, Touré-Tillery,

& Fishbach, 2020; Yamagishi, Terai, Kiyonari, Mifune, & Kanazawa, 2007).

Overview of Studies

We conducted three studies that test whether responsibility amplifies empathic forecasts (RAEF). In Study 1, participants were either personally responsible for selecting a positive and negative experience for another participant, or they knew one would be randomly chosen. We predicted that the prospect of being responsible for visiting these experiences upon another would amplify participants' own reactions to them and, in turn, empathic forecasts. Study 2 aimed to understand what features of social actions produce this amplification, by testing whether the combination of causality and intent does indeed amplify empathic forecasting. After reporting evidence that the self is more focused on its own causal responsibility for a looming decision than it is the responsibility of another intentional agent, we present Study 3; it examines whether actions for which the self (as opposed to another person) is responsible are those for which empathic forecasts are amplified. Data and materials for all studies are posted online: <https://osf.io/apbv5>.

Study 1

In Study 1, participants predicted others' reactions to a pleasant or aversive slideshow. Sometimes the self would be responsible for what the other would see. In other cases, mere randomness would be responsible for the outcome. We hypothesized that the slide-shows would be assumed to evoke more affect if the self had responsibility for the outcome.

In this and the subsequent studies, we include both positive and negative outcomes. This allows us to distinguish support for our hypothesis (that responsibility amplifies empathic forecasts) from the orthogonal possibility that the self as a responsible agent merely produces more positive or more negative feelings in others. In addition, we expected that participants would themselves experience stimuli more strongly when they would be responsible for the recipient's experience, and that this would statistically explain the effect of responsibility on empathic forecasts. It is not until the next study that we examine more closely what it is about the self's actions that creates such amplification, but the hypothesized mediation could provide initial support for our account.

Method

Participants and design. Lacking knowledge of the true effect size by which responsibility amplifies empathic forecasts (much less the true effect size in this paradigm), we did not know how many participants would be necessary to achieve adequate power. Instead, we took several steps to maximize power. First, we recruited participants from two samples simultaneously, given our access to both samples (an undergraduate subject pool at the University of California, Berkeley, and Amazon's Mechanical Turk [AMT]) and knowledge that we would achieve greater statistical power through the larger sample size that leaning on both participant populations permitted. Second, we wanted to exceed Simmons, Nelson, and Simonsohn's (2013) rough guideline that sample sizes under 50 participants per condition are suspect. We

recruited as many participants through the undergraduate subject pool as we had access to in one semester ($n = 164$) and as many participants through AMT as the funding lab's monthly budget—divided among all studies run—would permit ($n = 135$). Third, where feasible, we leaned on within-subjects manipulations that would allow us to increase statistical power (under the assumption that participants' multiple responses would be correlated). The relevant Institutional Review Boards approved all studies reported in this article.

We used a 2 (outcome responsibility: self or randomness) \times 2 (valence: positive or negative) mixed design. Only the first factor was manipulated between subjects. Of the 299 participants, 25 were unable to pass an attention check—a simple multiple-choice question asking what they had done in the study—and were excluded from all analyses. The exclusion rate did not vary by the between-subjects factor, outcome responsibility, $\chi^2 < 1$. The check itself as well as analyses without exclusions are reported in the [online supplemental materials](#).

Procedure. At the start of the experiment, participants were taken to a loading page where they were told they would wait up to 60 s, ostensibly to be paired with another participant. After 10 s of seeing a page-loading gif (to reinforce the cover story), participants were told they had been successfully matched with another participant. The two participants would supposedly complete two tasks in the opposite order. This explained why participants would be watching the key slideshows before the other participant and sometimes selecting which the other would watch.

We created four brief slideshows. Each contained six images from the International Affective Picture System. Each image was displayed for 3 s; thus, each presentation lasted 18 s. Two slideshows were positive in valence and were matched on thematic content and average arousal (Lang, Bradley, & Cuthbert, 1997): They contained pictures of baby animals, children, and people engaging in fun activities. The other two slideshows were negative in valence and were also matched on content and average arousal. They displayed children with disfigurements, images of deceased animals and people, and a person persecuting others.

Participants were told the yoked other (the recipient) would be taking part in a different condition, one in which they would watch two of the four slideshows—one positive and one negative. When the self was responsible, participants were led to believe we would enlist their help in selecting which positive video and which negative video the experimenter would present to the recipient. But when randomness was responsible, participants thought that a computer would randomly select which positive and which negative video the experimenter would have the recipient watch. Crucially, participants learned these procedures before the videos began.

We had participants watch the positive and negative videos in a counterbalanced order. The sequencing of the two slideshows within each pair was also counterbalanced. Following each video, participants estimated how the yoked recipient would feel after watching that video, if indeed that were the one picked. Participants indicated their forecasts on eight 101-point slider scales (from 0 = *not at all* to 100 = *most I have ever felt*) that measured negatively valenced emotions (*angry, disgusted, anxious, guilty*) or positively valenced emotions (*amused, happy, proud, loving*). In

addition, participants also reported their own emotional reactions to the videos using the same scales.

Results

We began by testing whether responsibility amplified empathic forecasts. More specifically, would forecasters think another would respond more strongly to a stimulus if it were one that might be intentionally selected by the self as opposed to randomly selected by a computer? For each video, we took the average of the valence-congruent emotional responses (positive emotions for positive slideshows, negative emotions for negative slideshows) and then subtracted the incongruent emotional responses (negative emotions for positive slideshows, positive emotions for negative slideshows). In this way, higher scores always reflected a stronger empathic forecast (*recipient emotional impact*) for a particular video. We followed the same procedure using participants' own self-reported reactions to the clips to develop analogous *self emotional impact* scores.

To understand whether the outcome responsibility manipulation changed participants' empathic forecasts and/or their own experience, we used 2 (outcome responsibility: self or randomness) × 2 (valence: positive or negative) mixed models, with only the second factor measured within subjects. One model predicted the recipient emotional impact scores; the other, the self emotional impact scores. Each model included two fixed effects—outcome responsibility (−1 = randomness, +1 = self) and valence (−1 = negative, +1 = positive)—as well as their interaction. The models also included three random effects to account for nonindependence. These were for participant, video (which of the four was being rated), and sample (Mechanical Turk or lab). Model estimates are provided in Table 1.

Recipient emotional impact. When the self (instead of a random algorithm) would decide what the recipient would view, participants displayed amplified empathic forecasts. That is, those considering which video to send to the other estimated the slideshows would evoke a stronger emotional response ($M = 47.78, SE = 0.97$) than those told a computer would make the assignment ($M = 41.89, SE = 0.98$), $F(1, 271.22) = 5.64, p = .018$ (see Figure 1). This offers initial support for RAEF. Unexpectedly, this effect was moderated by the valence of the video, $F(1, 817.11) =$

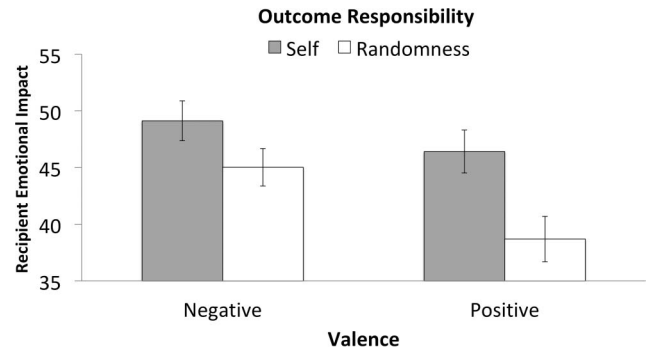


Figure 1. Study 1: Predicted recipient emotional impact as a function of the outcome responsibility manipulation and the stimulus valence. Higher values reflect a prediction that the recipient will respond more positively (vs. negatively) to the positive valence stimuli and more negatively (vs. positively) to the negative valence stimuli. All error bars reflect ±1 standard error of the mean. RAEF is reflected by a main effect of outcome responsibility.

3.93, $p = .048$. As shown in Table 1, this reflected that RAEF emerged more strongly for positive than for negative videos.

Self emotional impact. We tested whether responsibility amplified the self's own experience of the stimuli as well. Toward this end, we submitted the self's emotional impact scores to the same model. Participants who would be directly responsible for the other's outcome experienced the stimuli as more emotionally intense ($M = 45.30, SE = 1.07$) than those told the computer would decide ($M = 37.80, SE = 1.18$), $F(1, 271.35) = 7.61, p = .006$. This showed the same unanticipated moderation by valence, $F(1, 817.03) = 8.68, p = .003$.

Did deciders' own amplified experience predict their amplified empathic forecasts? We added the self's emotional impact scores to our initial mixed model. Consistent with the idea that self outcome responsibility participants' own elevated emotional experience was projected onto recipients, we observed a significant effect of self emotional impact, $t(940.84) = 49.25, p < .001$. With the self emotional impact controlled, RAEF disappeared, $t(262.59) = 0.296, p = .768$. A significant Sobel test provided evidence consistent with full mediation, $z = 2.75, p = .006$. In the online supplemental materials,

Table 1
Study 1: Effects of Outcome Responsibility, Valence, and Their Interaction on Recipient Emotional Impact (Empathic Forecasts) and Self Emotional Impact

Predictors	Recipient emotional impact (IVs → DV)	Self emotional impact (IVs → Mediator)	Recipient emotional impact (IVs, Mediator → DV)
Outcome Responsibility	2.61 (1.10)*	3.46 (1.25)**	0.14 (0.47)
Negative Valence	2.01 (1.21)†	2.13 (1.40)	
Positive Valence	3.26 (1.31)*	4.75 (1.58)**	
Valence	−2.27 (0.81)	−3.68 (0.55)*	0.34 (0.43)
Outcome Responsibility × Valence	0.91 (0.46)*	1.60 (0.54)**	−0.22 (0.29)
Self Emotional Impact	—	—	0.71 (0.01)***

Note. All values are unstandardized betas (and standard errors). The effects of negative and positive valence under outcome responsibility reflect the effect of outcome responsibility for the negative and positive valence outcomes, modeled separately. We include these simple effects for the two models in which the Outcome Responsibility × Valence interaction achieved statistical significance.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

we explain how we transformed our data in a way that would permit us to conduct a parallel test using Hayes's (2013) bootstrapping procedure. As described in full there, evidence in support of the indirect effect was robust to this analytic variant.

Did responsibility merely encourage more careful attention to the stimuli? We argued that the responsibility of the self—acting as a causal, intentional agent—for the outcome elevates the self's own emotional experience of the action and the outcome. But perhaps when the self would be responsible for another's fate, it may simply attend more carefully to those stimuli it might send to another. To test this possibility, we conducted a follow-up study that used a similar sample size ($N = 357$; $n = 146$ Americans from AMT, $n = 211$ UC-Berkeley undergraduates) and drew from the same populations as in the main study. But instead of examining emotional impact (recipient or self), we probed in two ways how carefully participants were attending to the slideshows.

After seeing all four slideshows, participants completed two self-report items: "How much attention were you paying while watching the slideshows?" and "How closely were you watching the slideshows?" Responses were provided on 7-point scales anchored at 1 (*Not much/Not closely at all*) and 7 (*A lot/Extremely closely*). These items were correlated ($r = .73$) and averaged. When the self was responsible for the recipient's outcome, it reported paying no more careful attention ($M = 6.02$, $SD = 1.09$) than when the outcome would be determined randomly ($M = 6.07$, $SD = 1.04$), $t < 1$.

Of course, self-reports may be distorted. Perhaps participants felt like they needed to assure the experimenter that they had paid careful attention. Thus, we also included a surprise memory test. Participants were asked to recall as many of the 24 images as they could. They were told to "identify each with enough specificity that someone else would know which image you are referring to." Participants had to spend at least 1 min on this task. A coder identified each response that was described in sufficient detail that he knew which image was being referred to. Once again, we found no evidence that those in the self outcome responsibility condition had attended more closely to the images. They recalled just as many pictures ($M = 7.22$, $SD = 4.64$) as did those who thought the other's videos would be randomly selected ($M = 7.26$, $SD = 4.91$), $t < 1$.

Twenty-six participants did not enter any recollections. As a result, these participants were coded as having zero correct recollections. We reconduted our analysis excluding these 26 participants, to make sure these maximally unengaged participants were not skewing our results. We still found that self outcome responsibility participants correctly recalled no more images ($M = 7.68$, $SD = 4.40$) than did randomness outcome responsibility participants ($M = 7.97$, $SD = 4.65$), $t < 1$.

Summary. When participants themselves would be responsible for what emotionally laden experience would be foisted upon another (vs. knew such outcomes would be imposed randomly), they experienced the stimuli as more emotionally impactful and estimated that the recipient would as well. That these two causal effects were correlated is statistically consistent with the possibility, but does not logically demand, that one effect was responsible for the other. But given many decades of demonstrations that the self projects its own experience onto others (e.g., Van Boven & Loewenstein, 2003), this causal pathway is theoretically plausible.

We experimentally varied the outcomes' valence so that the evidence in support of RAEF would not also support the possibility that responsibility merely elevates the positivity (or negativity) of empathic forecasts. The main effect of responsibility is thus inconsistent with this alternative possibility. That said, the fact that RAEF was observed more strongly for positive than negative outcomes was unanticipated. To foreshadow, we do see some variability across our studies in whether RAEF is moderated by valence, but those findings are not consistent. Given each study's positive and negative outcomes were not intended to be perfectly matched (e.g., Study 1's positive and negative videos varied in their content beyond just their valence), such inconsistent moderation may indicate that idiosyncratic features of certain positive or negative outcomes lend themselves more or less to RAEF.

Study 2

When we said in Study 1 that the self would be responsible for the recipient's outcomes, we meant that the self would both cause and intend for the recipient to experience one event as opposed to another. After all, this is what the self would be doing in selecting a slideshow for the recipient to see. But Study 1 did not provide direct support that these psychological features explained the amplified empathic forecasts. Study 2 aimed to test whether this combination of features—those that compose what we have called responsibility—amplify empathic forecasts.

In Study 2, we designed a set of six scenarios in which the self was asked to consider having some degree of responsibility for a recipient's positive or negative outcome. More specifically, for each scenario, we used a variety of tools that past research has shown affect perceptions of causality and/or intentionality. By doing so, we created parallel scenarios in which the self was directly responsible (meaning causality and intent were hypothesized to be high) or indirectly responsible (meaning causality and intent were hypothesized to be low). In this way, the self always considered being the agent, but we either amped up or reduced the crucial ingredients that we thought explained the degree of assumed affective response in the recipient.

Although causing and intending an outcome are theoretically distinguishable, previous research has found that it is not easy to cleanly manipulate one without manipulating the other (Cushman & Young, 2011). That is, judgments of causation are themselves influenced by judgments of intentionality (Shultz, 1980; Shultz & Wright, 1985). Furthermore, intention has been identified as a central factor in personal causality (Heider, 1958). This is why we created scenarios for which the self was directly (i.e., highly) or indirectly (i.e., not very) responsible for the recipients' outcomes, with the idea that 1) direct responsibility scenarios would generally be higher on both causality and intent, but also 2) these manipulations would produce some variability across scenarios and across people in the extent to which perceptions of causality and intent were elevated. Such variability would be important to allow us to test whether it is the special combination of cause and intent (or, in terms of statistical mediation, a positive interaction between these variables) that explains what about greater responsibility amplifies empathic forecasts.

Our scenarios used a mix of three experimental tools to vary the self's responsibility for a recipient's outcome. First, in some direct responsibility scenarios the self took a specific action to bring

about the outcome, whereas in the indirect responsibility version the self passively permitted the outcome to occur uninterrupted (Baron & Ritov, 2004; Cushman & Young, 2011; DeScioli, Bruening, & Kurzban, 2011; Spranca et al., 1991). Second, direct responsibility sometimes meant the self was clearly the source that instigated the action, whereas with indirect responsibility the action was originally instigated by a random act of nature (Ames & Fiske, 2013). Third, in some direct responsibility cases the self both desired the outcome and thought that through its own actions it could bring them about; in the indirect responsibility counterparts, this combination was lacking (Malle & Knobe, 1997). By leaning on various ways that responsibility has been manipulated, we hope our tests speak to the influence of responsibility on empathic forecasts more generally.

We had three predictions. First, we expected that our direct responsibility scenarios would produce amplified empathic forecasts. Second, as something of a check on our responsibility manipulations, we expected that the self would be seen as more clearly causing and intending the recipients' outcomes in the direct (vs. indirect) responsibility scenarios. Third, we expected that the elevation in both causality and intent would explain RAEF, but we thought the two components would amplify empathic forecasts most strongly in combination. In other words, it was important that we empirically distinguished cause and intent to determine whether each most clearly predicts RAEF when the other component is present as well.

Method

Participants and design. When we ran Study 2, we did not have access to a university subject pool. One thousand six hundred twenty-eight Americans were recruited from AMT and paid a nominal amount for their participation. We used a 2 (Valence: positive or negative) \times 2 (Responsibility: direct or indirect) mixed design. Only the first factor was varied between subjects. For each scenario, we randomly assigned each participant to see one of two Responsibility versions: one for which the self was characterized as having direct responsibility for another's outcome, one for which the self had indirect responsibility. Note that by using random assignment at the level of each scenario that a participant confronted (as opposed to merely at the participant level), the outcome responsibility factor typically varied within subjects (even though each participant saw only one version of each scenario). One hundred thirty-seven participants failed an attention check and were excluded from all subsequent analyses. This exclusion rate did not vary by the between-subjects factor, valence, $\chi^2 < 1$. We report analyses including the remaining 1,491 participants below. Analyses using these excluded responses as well are presented in the [online supplemental materials](#).

Procedure. To begin, participants learned they would consider six situations they could find themselves in. Although all participants considered scenarios that related to the same six situations—managing a workplace's employee review system, holding a coveted seat on a subway, being a football referee, grading exams, staffing a boot camp race, and assigning apartments to tenants—we modified each scenario so that the self's actions would ultimately relate to positive or negative outcomes for another. Depending on participants' valence condition, they

saw the positive or the negative version of all six vignettes. The six were presented in a randomized order.

For each of these 12 vignettes, we created two versions. In one version, the self's actions made him or her directly responsible for the target's emotionally valenced experience. In the other version, the self was expected to seem indirectly responsible for the other's experience. Given our interest in testing the effects of responsibility more generally instead of the influence of any one specific operationalization, the specific tools used to produce each direct and indirect responsibility scenario included some mix of whether the self directly acted (vs. failed to act), whether the self tried to bring about an outcome (vs. having it come about as a side effect of another outcome), and/or whether the self took actions that it believed would produce a desired outcome (vs. this combination not being present).

Consider the vignette in which the self considers refereeing an intramural football game. Each version began in the same way:

Your company is having an intramural flag football game. You're not much into playing, so you decide to take on a referee position for the game. You and another person are the only two referees, so you're both responsible for calling the game.

At that point, all participants learned that the trailing team has the ball and throws a pass, but that the receiver falls while trying to catch it. In all four versions, it then says that it is unclear from everyone else's angle whether the pass was complete.

In the *positive direct responsibility* version, the scenario ended with, "The other referee was turned around when the play happened, so he points to you to make the call. You know the pass was incomplete, but instead you decide to call it a touchdown. The team wins the game." In the *positive indirect responsibility* version, the scenario ended with, "The other referee calls it a touchdown. You know the pass was actually incomplete, but decide not to overrule the other referee's call. The team wins the game."

In the *negative direct responsibility* version, the scenario ended much as the positive direct responsibility version did (with the referee being turned around and pointing to you to make the call), except in this case, "You know it was a catch, but decide to call it incomplete. The team loses the game." In the *negative indirect responsibility* version, the vignette instead paralleled the positive indirect responsibility version, except this time the referee called the pass incomplete. The scenario then concluded with, "You know it was a catch, but decide not to overrule the other referee's call. The team loses the game."

Note how in every version the self serves as a referee who has some role in causing the team to win (positive outcome) or lose (negative outcome) the game. But the self achieves these outcomes through an act of commission (making the call) or an act of omission (failing to overrule the other referee's call). Also, in every version, we make clear that the self knows that the final call is the wrong one. All that we vary is whether the self played an active or passive role in bringing about the outcome.

We wish to call attention to a few features that guided the construction of these vignettes. Although in Study 1 we did not tell participants that recipients would know that the videos had been selected by another participant or a computer, we also did not tell them that recipients would *not* know this. In the referee scenario, like in all the vignettes, we were careful that the responsibility

manipulation would not change the nature of the recipient's experience or understanding of why the outcome occurred. (The win or loss was always determined by a wrong call from one of the two referees, and it is not shared that the self knowingly makes or permits the wrong call.) Where possible, the positive and negative versions were parallel (e.g., winning vs. losing the football game). The outcome valence factor should be thought of as a way to assess the robustness of RAEF as opposed to a precise method for determining whether RAEF differs in size for positive and negative outcomes. Table 2 summarizes the crucial features of each scenario, which are presented in full in the [online supplemental materials](#).

Following each vignette, participants responded to three measures. The first two, presented in a counterbalanced order, asked the extent to which the self caused and intended the outcome. The cause item asked, "How much of a role did you play in causing [recipient outcome]?" the intent item read, "To what extent did you intend for [recipient outcome]?" Each response was offered on a 9-point scale anchored at 1 (*Not at all*) and 9 (*Very much*). The third measure was for the recipient emotional impact. For positively valenced vignettes, it read, "Estimate how happy [recipient] will feel after [outcome]." For negatively valenced vignettes, "happy" was replaced with "irritated." Participants responded on 101-point slider scales anchored at 0 (*Not at all*) and 100 (*Most they have ever felt*).

Results and Discussion

Recipient emotional impact. To begin, we tested whether placing the self in a directly (vs. indirectly) responsible role led to stronger forecasts of recipient emotional impact. We constructed a mixed model predicting the emotional impact scores. We included several fixed effects: valence ($-1 = \text{negative}$, $+1 = \text{positive}$), responsibility ($-1 = \text{indirect}$, $+1 = \text{direct}$), and their interaction. We also included random effects of *scenario* and *participant* to account for nonindependence among certain responses.

First, we observed a main effect of responsibility, $B = 0.84$, $SE = 0.16$, $t(7989.05) = 5.14$, $p < .001$. This provided direct support for RAEF: As the self's direct responsibility for the recipient's outcome increased, participants offered elevated empathic forecasts. As presented in the [online supplemental materials](#), this pattern achieved statistical significance in four of the six scenarios ($ps < .05$) when analyzed individually. A significant Valence \times Responsibility interaction emerged as well, $B = -0.41$, $SE = 0.16$, $t(7988.80) = 2.53$, $p = .011$. Whereas in Study 1 responsibility amplified empathic forecasts somewhat more for positive compared to negative outcomes, in this case responsibility amplified empathic forecasts somewhat more for negative than positive outcomes. Decomposing the data showed that responsibility amplified empathic forecasts for both negative outcomes, $B = 1.25$, $SE = 0.23$, $t(4111.15) = 5.34$, $p < .001$, and positive outcomes, $B = 0.48$, $SE = 0.22$, $t(3848.65) = 2.22$, $p = .026$. We next turn to what it is about the responsibility manipulation that explains the amplified empathic forecasts.

Decomposing responsibility: Cause and intent. We next tested whether the responsibility manipulation did indeed affect how much the self was seen to cause and intend the outcomes described in the vignettes. When the self was directly responsible for the recipient's fate, the self was seen to have caused that

outcome more, $B = 0.93$, $SE = 0.02$, $t(8564.40) = 43.95$, $p < .001$. This was actually truer for the positive than the negative events, $B = 0.25$, $SE = 0.02$, $t(8564.03) = 11.73$, $p < .001$. But crucially, the effect on causality emerged both for participants considering the positive, $B = 1.17$, $SE = 0.03$, $t(4233.70) = 37.28$, $p < .001$, and those considering the negative scenarios, $B = 0.67$, $SE = 0.03$, $t(4395.46) = 21.86$, $p < .001$. Being directly responsible for the recipient's fate also elevated judgments that such outcomes were intended, $B = 1.48$, $SE = 0.03$, $t(8534.92) = 58.08$, $p < .001$. In this case, the effect did not differ between positive and negative outcomes, $t < 1$.

We proceeded to ask whether the extent to which the self was responsible for an outcome—meaning the self both caused and intended the recipient's end result—would explain the boost in empathic forecasts. Toward this end, we returned to our original model but added in three terms: cause and intent (both standardized) as well as their interaction. Both cause, $B = 2.41$, $SE = 0.26$, $t(8513.82) = 9.35$, $p < .001$, and intent, $B = 2.01$, $SE = 0.23$, $t(8391.63) = 8.90$, $p < .001$, independently predicted the amplified empathic forecasts. But speaking to the particular power that cause and intent combine to have, we observed a significant Cause \times Intent interaction, $B = 0.81$, $SE = 0.19$, $t(8371.71) = 4.35$, $p < .001$ (see Figure 2).

Especially after seeing that valence moderated some of our earlier effects, we proceeded by decomposing this interaction separately for positive and negative vignettes, at plus or minus one standard deviation on each predictor. The picture was largely consistent (see Table 3). When the self was seen to play more of a role in causing the recipient's outcome ($+1$ standard deviation), more intent predicted greater empathic forecasts. And when the self was seen to more clearly intend the recipient's outcome ($+1$ standard deviation), the self's causal role predicted greater empathic forecasts. This was true for both the positive and negative outcomes. In contrast, when the self was seen to play less of a role in causing (-1 standard deviation) or showed less evidence of intending (-1 standard deviation) the outcome, the other factor tended not to significantly account for variability in empathic forecasts. This suggests that cause and intent are best thought of not as independent mediators, but as mutually reinforcing contributors to RAEF. In other words, it is the combination of cause and intent that links the self as a responsible agent to a highly experiential recipient.

Study 3

Study 3 extended on our previous two studies in three main ways. The first was reflected in a new condition. In our previous studies, the causal intentional agent either was (Study 1) or was said to be (Study 2) the self. Study 3 added a new *other* outcome responsibility condition in which the intentional agent that would affect another's fate was another person.

Second, we used a new context: a dictator game. Participants were led to believe that they were yoked to one or two other participants. They learned that they themselves (*self* condition), one of the other participants (*other* condition), or the random determination of a computer (*randomness* condition) would determine how much of \$10 would be allocated to someone else. Third, we went to even greater lengths than in Study 2 to make certain that the objective experience of the recipient would be equated

Table 2
Study 2: Summary of Scenarios, Including Manipulations of Direct and Indirect Responsibility and Outcome Valence

Context	Agent	Recipient	Positive outcomes			Negative outcomes		
			Direct	Indirect	Recipient outcome	Direct	Indirect	Recipient outcome
Employee review	Manager	Employee	Manager alters the employee's ratings	Manager fails to correct an error in the employee's ratings	Employee receives a full bonus	Manager alters the employee's ratings	Manager fails to correct an error in the employee's ratings	Employee receives only half their bonus
Subway	Train passenger	Elderly man	Train passenger changes cars, knowing the elderly man is hoping to sit	Train passenger exits car because it is their stop	Elderly man takes the passenger's seat	Train passenger leaves their luggage on a seat, knowing the elderly man is hoping to sit	Train passenger leaves their luggage on a seat, not noticing the elderly man is waiting to sit	Elderly man must stand for the duration of his trip
Football game	Referee	Player	Referee knowingly makes an incorrect call	Referee fails to overturn an incorrect call	Player wins the game for their team	Referee knowingly makes an incorrect call	Referee fails to overturn an incorrect call	Player loses the game for their team
Exam grading	Teaching Assistant	Student	Teaching Assistant uses a known-to-be-faulty Scantron machine	Assistant uses a Scantron machine that surprisingly malfunctions	Student receives an 'A' on an exam	Teaching Assistant uses a known-to-be-faulty Scantron machine	Teaching Assistant uses a Scantron machine that surprisingly malfunctions	Student fails an exam
Boot camp	Staffer	Boot camp participant	Staffer rigs a supposedly random process that assigns tasks	Staffer mistakenly allows a task assignment to be truly random	Participant bypasses an unpleasant task	Staffer rigs a supposedly random process that assigns tasks	Staffer mistakenly allows a task assignment to be truly random	Participant falls into mud
Apartment rental	Landlord	Tenant	Landlord overrides the computer lottery system	Landlord fails to correct computer error in the lottery system	Tenant moves into an especially luxurious apartment	Landlord overrides the computer lottery system	Landlord fails to correct computer error in the lottery system	Tenant moves into a hazardous apartment where a pipe burst ruins their things

Note. Participants saw only one version—defined by Responsibility and Outcome Valence—in each row. Participants always saw versions of only one Outcome Valence, but which Responsibility version they saw was randomly assigned for each scenario for each participant.

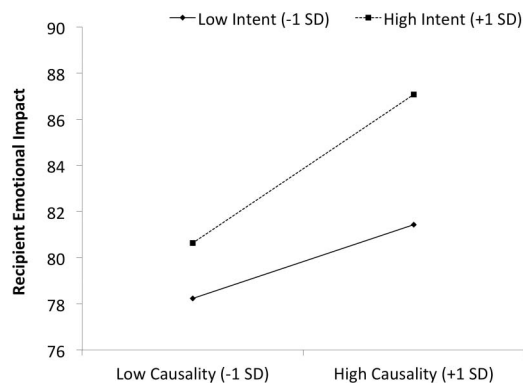


Figure 2. Study 2: Recipient emotional impact scores, predicted at low (−1 SD) and high (+1 SD) levels of intent and causality. Higher values reflect amplified empathic forecasts. The significant interaction reflects that RAEF emerges when both high causality and high intentions are dually perceived.

across conditions. In all conditions, the recipient would supposedly be told that the amount of their allocation had been randomly determined by a computer (even when it was actually the self or another person who would make the determination).

We asked participants to predict how the recipient would respond to four different allocations. To determine whether RAEF replicated, we tested whether participants would forecast that recipients would show more of a differentiated emotional response to these allocations. Note we now had four orderable levels of outcome valence (instead of just two as in the earlier studies). As such, we tested whether the recipient was assumed to be more pleased as they received increasing amounts when the self as opposed to pure randomness was responsible for the outcome. Of greater interest was whether allocations that were caused and determined by another intentional agent (i.e., someone other than the self) would be seen as especially emotionally impactful as well. If causal responsibility is what leads agents to be seen as having a strong emotional impact on their recipients, then a key question is whether forecasters similarly focus on and thus feel the weight of such responsibility regardless of whether they themselves or someone else is confronting that choice. After all, the self is typically held responsible for its own (and not others’) choices, so such greater attention to the self’s causal responsibility would seem natural. For those who have ever lent a sympathetic ear to someone

confronting a difficult social decision, it may be easy to intuit that someone else’s looming responsibility does not weigh as heavy phenomenologically as when the self faces such a choice.

With this in mind, we conducted a pilot study asking whether self and other outcome responsibility condition participants in Study 3 are likely to be equally focused on the causal agent’s responsibility for the recipient’s outcome. Americans recruited from AMT ($N = 455$) took part in a study modeled after the main study (whose full methods are described below), but in which all participants were assigned to the *self* or *other* outcome responsibility condition. Instead of asking participants to forecast the recipient’s emotional response to different allocations, we asked the self [other] participants, “As you were preparing to make that decision [learn what decision the other person would make], how much were you thinking about. . . .” Participants responded to three prompts on a scale from 1 (*not at all*) to 7 (*a lot*). One is relevant for the present purpose: “the fact that you [the other participant] would cause (i.e., be responsible for) how much money someone else would get.” (See the [online supplemental materials](#) for the wording of the other two items, as well as null between-conditions effects on them.) We found that participants were egocentrically focused on their own responsibility ($M = 5.62, SD = 1.56$) instead of that of the other decider’s ($M = 5.23, SD = 1.55$), $t(453) = 2.57, p = .011, d = .24$. This egocentric focus on one’s own responsibility suggests that at least in this paradigm RAEF may be egocentric as well: The self may assume it will have a bigger emotional impact on another than would another intentional agent.

Method

Participants and design. Guided by the same rules laid out in Study 1, we sought to achieve a large sample size by recruiting participants simultaneously from AMT ($n = 281$) and an undergraduate subject pool ($n = 307$). AMT participants were Americans who were paid a nominal amount. Undergraduate participants received course credit at the New York University. These 588 participants were randomly assigned to one of three outcome responsibility conditions: self, other, or randomness.

Procedure. As in Study 1, we began by leading participants to a loading page where they were to wait up to 60 s while they were ostensibly paired with one (*self* and *randomness* conditions) or two (*other* condition) other participants. After 10 s, participants were told that this pairing had occurred. All participants then learned

Table 3
Study 2: A Decomposition of the Interactive Effects of Cause and Intent on Predicted Recipient Emotional Impact by (Outcome) Valence

Simple slope	High (+1 SD)	Low (−1 SD)
Positive Outcome Valence		
Effect of Causality when Intent is. . .	B = 1.01 (0.48)*	B = 0.17 (0.30)
Effect of Intent when Causality is. . .	B = 1.38 (0.39)***	B = 0.54 (0.37)
Negative Outcome Valence		
Effect of Causality when Intent is. . .	B = 8.16 (0.69)***	B = 3.60 (0.31)***
Effect of Intent when Causality is. . .	B = 3.79 (0.41)***	B = −0.76 (0.51)

Note. Each B (and standard error) reflects the effect of causality or intent on recipient emotional impact when the other predictor is relatively high (1 SD above the mean) or low (1 SD below the mean).
* $p < .05$. *** $p < .001$.

about a \$10 pot that would be split between two people. How that split would be determined and which two participants would receive the money varied by condition.

Those in the *self* outcome responsibility condition learned they (as a decider) would determine how much of the money to give another participant (the recipient), thereby leaving the remainder for themselves. Those in the *randomness* condition were also told that the \$10 would be split between themselves and another participant (the recipient), but learned that a computer would randomly determine the allocation. Those in the *other* condition—who learned they had been yoked to two others, a decider and a recipient—were told the decider was facing the same decision that those in the self outcome responsibility condition were: a decision of how much of the \$10 to give the recipient and how much to keep for himself or herself. In order to avoid confounding the outcome responsibility manipulation with the self's own financial interest in seeing the target receive less money, we told participants in the other outcome responsibility that their own financial outcome (merely as an observer of the other's decision) would be yoked to what the decider kept. In this way, forecasters (regardless of condition) always stood to gain or suffer from the recipient's loss or windfall, respectively.

In all conditions, participants considered how the recipient would feel if that person were to receive a specified amount of the \$10: \$10, \$5, \$1, or \$0.01. Regardless of condition, the recipient was supposedly informed that their allocation would be determined randomly by a computer. For each amount, participants provided an estimate of the recipient's emotional reaction on eight dimensions: contented, pleased, satisfied, happy, cheated, disappointed, upset, and frustrated. Each judgment was made on a 101-point scale ranging from 0 (*not at all*) to 100 (*most they have ever felt*).

In Studies 1 and 2, we had outcomes that could be dichotomized as positive or negative. For that reason, we always coded our recipient emotional impact scores so that higher values would reflect stronger empathic forecasts in the expected valenced direction. In the present study, outcomes instead exist at four levels along a continuum from most negative to most positive. Because of this methodological difference, we always coded the recipient emotional impact scores in the same way, so that we could examine differences between conditions in this index's sensitivity to changes in allocation. That is, we ask, as the allocation grew, when did participants assume that the recipient's (positive) response would most grow as well? After reverse scoring the final four (negative) emotional reactions, we averaged the items to create a recipient (positive) emotional impact score ($\alpha = .87$).

Results

We sought to test whether allocations' perceived ability to influence the recipient's emotional impact depended on who or what would decide that allocation. Toward that end we created a new variable, allocation. This reflected the monetary value received by the recipient: \$10 (+3), \$5 (+1), \$1 (-1), and \$0.01 (-3). Although we are not endorsing the idea that the empathic forecasts across these levels would be perfectly linear, we used these codes to capture the hypothesized monotonic relationship.

We conducted a mixed model in which we included fixed effects of allocation and outcome responsibility condition (a cat-

egorical variable reflecting self, other, or computer), as well as their interaction. To account for nonindependence, we included random effects of participant and sample (Mechanical Turk or lab). Unsurprisingly, we observed a strong effect of allocation on recipient emotional impact, $F(1, 2345) = 2,915.24, p < .001$. This reflected that participants assumed that recipients would grow more pleased as their allocation increased. But most central to our hypotheses, an Outcome Responsibility \times Allocation interaction emerged, $F(2, 2345) = 15.92, p < .001$. Before decomposing this interaction, we also note there was no effect of outcome responsibility, $F < 1$. Because of the way the model was specified, this null effect suggests that any evidence of RAEF—unlike in Studies 1 and 2, in which positive and negative outcomes, respectively, showed more evidence of RAEF—did not vary by outcome valence.

We proceed to decompose this interaction into a series of 2 (Outcome Responsibility) \times Allocation interactions. The mean recipient emotional impact scores by condition for each allocation level are presented in Figure 3. In the text, we will reference a difference score that offers a simple picture of how much emotional impact an agent is seen to have: the target's predicted emotional response to receiving \$10 (the highest outcome) minus the target's predicted emotional response to receiving 1 cent (the lowest outcome). In what follows, we call this value M_{dif} . First, a significant 2 (Outcome Responsibility: self or randomness) \times Allocation interaction, $t(2345) = 5.44, p < .001$, provided initial support for RAEF. That is, participants thought allocations would have a larger impact on recipients' emotional state if participants themselves ($M_{dif} = 72.15, SD = 32.96$), as opposed to randomness ($M_{dif} = 55.54, SD = 49.32$), were responsible.

Recall that pretest participants who would be personally responsible for the target's fate (self outcome responsibility condition) reported being more focused on their own responsibility for the target's fate than other outcome responsibility participants were focused on someone else's responsibility for that same fate. As foreshadowed by the pretest, we also observed a significant 2 (Outcome Responsibility: self or other) \times Allocation interaction, $t(2345) = 4.05, p < .001$. In other words, it was not that any intentional agent was assumed to have the same impact on another's emotional response. Instead, the self thought its own allocation decisions would more strongly determine recipients' states than would another's decisions ($M_{dif} = 60.09, SD = 46.97$).

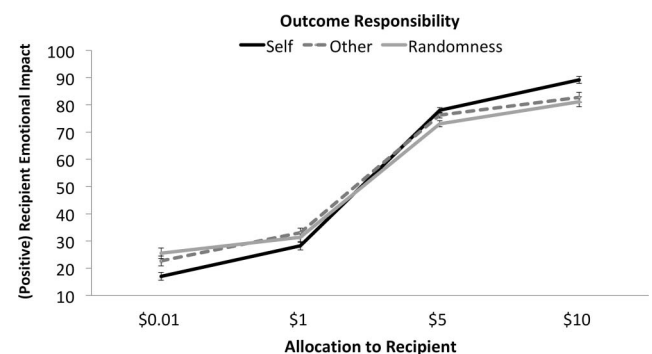


Figure 3. Study 3: Predicted positive recipient emotional impact as a function of the allocation to the recipient and the outcome responsibility manipulation. All error bars reflect ± 1 standard error of the mean.

We did not observe a significant 2 (Outcome responsibility: other or computer) \times Allocation interaction, $t(2345) = 1.41, p = .158$. This shows it is not always the case that intentional agents are assumed to amplify another's emotional response. Study 2 found that the combination of cause and intent elevated the assumed response in another. Study 3—combined with its pretest—found that forecasters may not be equally focused on and influenced by all agents' responsibility for outcomes. For the self, such responsibility looms large, and RAEF emerges. How this logic suggests that RAEF itself is not necessarily egocentric is a topic we return to—including with preliminary data—in the General Discussion.

General Discussion

Three studies examined the perceived power of responsibility, finding that responsibility amplifies empathic forecasts (RAEF). Study 1 showed that when participants (as opposed to the random determination of a computer) were responsible for which stimuli the experimenter would show another, participants estimated that the recipients would respond more strongly to them. When participants would be responsible for another's experience, they did not show evidence of attending more carefully to the stimuli (e.g., memory for them was not elevated). But those personally responsible for the outcome did find the stimuli more emotionally evocative for themselves as well. After all, generous parents may feel particular delight as they await their children opening presents on Christmas morning, even as they know that Santa will receive the credit.

Study 2 pinpointed what it is about agents' actions that amplify assumed responses in others. We had participants consider taking various actions that would directly (because the agent clearly caused and intended them) or indirectly (because the agent's responsibility was more ambiguous) affect another. Even though such targets would be blind to these forces, participants estimated that responsibility would amplify assumed responses. This was particularly true when one would be both causing and intending such outcomes.

Of course, it is not only the self that has the potential to cause and intend outcomes in others. But the self may be particularly prone to contemplate the reality of its own responsibility. As one contemplates ending a relationship, the weight of being potentially responsible for another's sadness almost certainly looms larger in the self's mind than in the minds of those with whom such potential plans have been discussed. And indeed, a pretest to Study 3 confirmed this intuition: When the self approached an economic game with the decision of how much money to allocate to another (as opposed to as an observer to such an allocation), the self gave more consideration to its own (as opposed to the other's) responsibility for a recipient's financial outcome. In light of Study 2's findings, it was then not surprising that only the self (but not another intentional agent) was assumed to have a bigger emotional impact on the recipient than would the random determination of a computer.

Need RAEF Be Egocentric?

In combination, these findings might seem to reinforce a well-supported theme in psychology: The self sees itself as playing a

special and outsized role in the social world. People overestimate their capabilities (Kruger & Dunning, 1999) and what they have control over (Thompson, 1999). The self not only exaggerates how much others notice them (Gilovich, Medvec, & Savitsky, 2000), but it also inflates its own contributions to group tasks (Robins & Beer, 2001). Furthermore, the self sees itself as highly efficacious, with lofty intentions that—more than its peers'—are likely to be realized (Helzer & Dunning, 2012).

On the other hand, some of our theoretical motivation might seem to call into question whether RAEF is necessarily an egocentric phenomenon. People judge others' actions by placing themselves in the others' shoes—what Miller and Cushman (2013) call *evaluative simulation*—which partly explains why people experience aversion at the thought of others' proximal, up-close-and-personal moral transgressions (Greene et al., 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Furthermore, we have characterized RAEF as extending a core insight that underlies work on moral typecasting (Gray et al., 2012). According to that tradition, moral transgressions are understood through the template of the moral dyad—an agent who causes intentional harm in a suffering patient. RAEF argues that this picture applies beyond moral misdeeds to more general assumptions about the emotional impacts of agents who cause and intend the outcomes they visit upon another. But moral typecasting was developed to understand basic properties of social cognition, not how the self makes sense of its own immoral actions. Should RAEF not apply more broadly as well?

Our logic does not demand that RAEF applies only to the self's own actions. But we did find in a pretest to Study 3 that—at least in that situation—the self is more keenly aware of and attentive to its own (as opposed to another's) role as being causally responsible for others' outcomes. But by this line of reasoning, if we were to find a context in which there was not such an egocentric focus on the self's responsibility, we should find that RAEF applies to social judgments as well.

Consider Study 2. Although participants were asked to picture themselves engaging in various actions for which they would have direct or indirect responsibility for another's outcome, the fact that no real decision was looming made us wonder how necessary it was that the self was the focal agent in these scenarios. After all, presenting a nearly equivalent description of another engaging in these actions would seem to make causal responsibility equally clear without any obvious features that would make that other's responsibility fade into the mind's background. Of course, this is an empirical question. In what follows, we present preliminary data to test it.

Pretest. We returned to the negative scenarios used in Study 2. We both retained our original ones in which the self was the agent and created modified versions in which we described the agent as another person. We conducted a pretest ($N = 482$ Americans from AMT) that was parallel to the one used for Study 3. Thirty participants who failed an attention check (see [online supplemental materials](#)) were excluded from the remaining analyses. Unlike when participants were preparing to play or watch a dictator game (Study 3 pretest), we no longer found an egocentric focus: There was no difference in attention to responsibility when participants considered the self ($M = 5.83, SD = 1.49$) as opposed to another ($M = 5.91, SD = 1.31$) performing each action, $t(450) = 0.66, p = .511, d = -.06$. Although cross-study com-

parisons should always be greeted with caution, both means were high, much like for the self condition in the Study 3 pretest. If RAEF is not necessarily egocentric (but the focus on responsibility, the key contributor to RAEF, sometimes is), this pretest suggests this paradigm may be a context in which we should not only be able to replicate RAEF, but find that it does not depend on whether the agent is the self or another person.

Main study. In the main study, participants ($N = 327$ Americans from AMT) took part in a 2 (Agent: self vs. other) \times 2 (Responsibility: direct vs. indirect) mixed design, with only the first factor manipulated between subjects. Thirty-one participants failed an attention check (see [online supplemental materials](#)). We report analyses including the remaining 296 participants below. Whether each participant saw the direct or indirect responsibility version of each scenario was randomly assigned for each scenario for each participant. Like in Study 2, participants judged how irritated the recipient would be (despite recipients' ignorance of the agent's level of responsibility for the outcome). Replicating Study 2, participants thought that directly responsible agents would produce greater irritation than the indirectly responsible ones, $t(1578.90) = 2.83, p = .005$. Furthermore, the responsibility manipulation successfully enhanced perceptions of how much the agent caused, $t(1651.63) = 14.40, p < .001$, and intended, $t(1652.99) = 23.65, p < .001$, each outcome. In addition, it was the positive interaction of cause and intent predicting empathic forecasts, $t(1750.77) = 5.76, p < .001$, that illustrated the importance of cause plus intent in producing RAEF. But crucially, we did not find that RAEF was stronger (and in fact trended weaker, although not significantly) when the self was the agent as opposed to another, $t(1578.68) = -1.62, p = .104$. Our pretests would suggest that sometimes people are quite focused on their own (at the expense of another's) agency (Study 3) and sometimes they are not (the present study). We do not explain precisely when this divergence does or does not emerge, but we return to this question below.

Remaining Questions

Our studies have established that responsibility amplifies empathic forecasts (RAEF), identified cause and intent as dually necessary contributors to this effect, and offered evidence that RAEF is not necessarily an egocentric phenomenon but one that may be associated with the extent to which one is keenly focused on an agent's responsibility for a target's outcome. That said, we see these as only initial efforts in understanding RAEF phenomena. As such, we outline four general questions that we see as next steps in advancing this line of research.

When is the focus on responsibility egocentric? Although we have identified the combination of an agent's cause and intent as the key factors guiding RAEF, we have not developed an account of when it is that the self is egocentrically focused on its own responsibility more than that of another agent's. One possibility is that when the self or another is confronting an actual decision (like in Study 3), the weight of one's own looming decision makes the focus on responsibility egocentric. When considering hypothetical actions that have been said to occur (like in the follow-up study reported in the *Need RAEF Be Egocentric?*), the responsibility that any agent played may come into clearer and more even focus. Understanding what affects when agents' causal

responsibility weighs more or less heavily on forecasters' minds will be important to better understand when and for which agents responsibility amplifies empathic forecasts.

Do any factors that influence perceived responsibility affect empathic forecasts? A number of factors are known to influence perceptions of agents' causal role and intent. Although we found that the combination of cause and intent amplifies empathic forecasts, whether any factor that affects these dual mediators will have parallel effects on empathic forecasts is unclear. Several programs of research have identified what could be called non-normative, or at least surprising, influences on judgments of cause and intent.

After all, causality itself is a psychological construction (Hume, 1740), such that even judgments of the self's own causal role can be mistaken or at least misremembered (Preston & Wegner, 2007; Wegner, 2002). Moreover, judgments of causality are themselves influenced by judgments of blame: A speeding driver was seen more to have caused an accident if he was rushing home to hide drugs, as opposed to an anniversary present, from his parents (Alicke, 2000). Knobe (2003) found that a CEO focused on profits and nothing else is ascribed intentionality for hurting but not helping the environment as a side effect of his business decisions. That is, judgments of intentionality are themselves sensitive to apparent indifference to what Hindriks, Douven, and Singmann (2016) called normative reasons, factors that merit attention but whose neglect inspires blame (Hindriks, 2008).

In these cases, judgments of cause and intent seem to be means to pile on and thus express more moral disapproval, not dispassionate analyses of an agent's causal role and specific motivations. As such, these enhanced perceptions of cause and intent may not be of the sort that—for example—draw agents closer to targets, thereby producing more emotionally evocative and thus seemingly impactful actions. On the other hand, the desire to blame (and, through possible analogous extension, praise) that these factors invite may also prompt forecasters to exaggerate the emotional impacts of such actions. These possibilities also await more direct test.

Is RAEF a bias? We have been careful to avoid claiming that responsible agents' emotional impact on others is overestimated. More generally, it may seem that we have sidestepped the question of where bias resides with respect to RAEF. Our intent is not to evade. It is simply that in considering whether RAEF is a bias, there are two ways one can consider this question:

A mistaken reliance on responsibility. One is whether it is inaccurate to assume that directly responsible agents have more emotional impact than those who operate unintentionally or without clear causal responsibility. By this standard, we believe we have established a bias. Although those who knowingly suffer at the hands of another (as opposed to on account of an unintentional agent) may experience more pain (e.g., Gray & Wegner, 2008), we were careful to equate targets' knowledge and experience in Studies 2 and 3. If people intuit that intentional transgressions produce more sting, then RAEF may be one of many biases documented in psychology that reflect an overgeneralization of a true lay belief (see Hoffman, Yoelli, & Navarrete, 2016).

Our scenarios (Study 2) and experimental cover story (Study 3) were written to keep targets' experience constant. But might our findings have been driven by a subset of participants misinterpreting that targets were actually aware of the agents' responsibility? For example, even though Study 3 participants were always told that the targets would think their monetary allocation was deter-

mined by a computer, perhaps some participants missed this explanation. If so, the amplified empathic forecasts of those in the self outcome responsibility condition may have come from this inattentive subset of participants.

By this account, we should have seen greater variability in empathic forecasts when the self's responsibility was high versus low. This is because the two subsets of participants in these conditions—that is, those who did versus did not assume the target was aware of the agent's responsibility—would show different patterns of results and thus produce more variability in the condition overall. But a reanalysis of Studies 2 and 3's data found that was not the case. In fact, the standard deviation of empathic forecasts was directionally (although, unsurprisingly, not necessarily significantly) smaller in the direct-responsibility scenarios (Study 2) and the self outcome responsibility condition (Study 3) than in the other conditions.

Mean-level biases in forecasting. But there is a second form of the accuracy question that will not give rise to an unqualified, generalizable answer: Do forecasters exaggerate the emotional impact of responsible agents or underestimate the effect of random or unintended outcomes? Given there is likely to be great variability in whether particular actions' emotional impacts are generally overestimated or underestimated, this means RAEF can either exaggerate bias (when such impacts are otherwise overestimated) or reduce bias (when such impacts are otherwise underestimated). To illustrate this point, we next consider a recent finding that relates to gift giving.

Cavanaugh, Gino, and Fitzsimons (2015) found that gift givers overestimate how much acquaintances will be delighted by a socially responsible gift (e.g., a donation to a charity in the acquaintance's name). According to RAEF, someone who accidentally made the donation in another's name may assume it will fill the other with less appreciation, which would (inadvertently) draw his or her forecast closer to accuracy. But Cavanaugh et al. (2015) also found that this overestimation bias disappears when estimating how a close friend would react to the intentional provision of the same socially responsible gift. In that case, RAEF would suggest that the accidental contributor would be the biased one, likely to underestimate the friend's appreciative feelings. As this brief discussion suggests, even if we were to determine which forecasts were closer to being accurate with regard to the impact of a particular action in one of our studies, any such conclusion would speak more to idiosyncrasies of that behavior rather than to the properties of RAEF itself. For this reason, instead of asking whether RAEF encourages or reduces forecasting bias, we see more value in encouraging the development of a more general account of when empathic forecasts tend to err toward overestimating or underappreciating emotional impact. With that knowledge, one would be better able to predict when RAEF pushes people toward or away from bias of this variety.

Might RAEF be functionally adaptive? Finally, there remain questions about whether RAEF is in fact functional in encouraging good deeds and discouraging bad ones. Anticipating clear joy or pain that could come from one's good or bad actions, respectively, may encourage the former but discourage the latter. On the other hand, RAEF may discourage people from performing the same quantity of good as they would have otherwise. That is, it may lead people to feel that even a relatively small donation or act of kindness will bring others a sufficiently desirable amount of happiness. Furthermore, people may learn to strategically sidestep

the immorality-discouraging constraints RAEF places on their antisocial impulses. After all, people have been shown to delegate to others the task of carrying out harms (Steffel, Williams, & Perrmann-Graham, 2016). It may be out of sight, out of mind: By being less focused on someone else being responsible for a third party's pain, the outcomes may seem less severe. Or those who carry out one's own desired misdeeds may be seen as less causally responsible for another's fate (Woolfolk, Doris, & Darley, 2006), meaning the consequences of their actions may be perceived as less severe. And when sidestepping direct responsibility is difficult, the prospect of RAEF may explain why people resort to dehumanizing their victims in order to counteract the inflated perceptions of pain that their own responsibility induces (Leyens et al., 2000). Ultimately, more work is necessary to determine how RAEF can be harnessed to encourage prosociality.

References

- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*, 556–574. <http://dx.doi.org/10.1037/0033-2909.126.4.556>
- Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science*, *24*, 1755–1762. <http://dx.doi.org/10.1177/0956797613480507>
- Ames, D. L., & Fiske, S. T. (2015). Perceived intent motivates people to magnify observed harms. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 3599–3605. <http://dx.doi.org/10.1073/pnas.1501592112>
- Arditte Hall, K. A., Joormann, J., Siemer, M., & Timpano, K. R. (2018). The impact bias in self and others: Affective and empathic forecasting in individuals with social anxiety. *Behaviour Research and Therapy*, *106*, 37–46. <http://dx.doi.org/10.1016/j.brat.2018.05.001>
- Baron, J. (1992). The effect of normative beliefs on anticipated emotions. *Journal of Personality and Social Psychology*, *63*, 320–330. <http://dx.doi.org/10.1037/0022-3514.63.2.320>
- Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, *94*, 74–85. <http://dx.doi.org/10.1016/j.obhdp.2004.03.003>
- Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*, 381–417. <http://dx.doi.org/10.1016/j.cognition.2008.03.001>
- Cavanaugh, L. A., Gino, F., & Fitzsimons, G. J. (2015). When doing good is bad in gift giving: Mis-predicting appreciation of socially responsible gifts. *Organizational Behavior and Human Decision Processes*, *131*, 178–189. <http://dx.doi.org/10.1016/j.obhdp.2015.07.002>
- Critcher, C. R., Dunning, D., & Rom, S. C. (2015). Causal trait theories: A new form of person knowledge that explains egocentric pattern projection. *Journal of Personality and Social Psychology*, *108*, 400–416. <http://dx.doi.org/10.1037/pspa0000019>
- Cushman, F. (2015). From moral concern to moral constraint. *Current Opinion in Behavioral Sciences*, *3*, 58–62. <http://dx.doi.org/10.1016/j.cobeha.2015.01.006>
- Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*, 2–7. <http://dx.doi.org/10.1037/a0025071>
- Cushman, F., & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, *35*, 1052–1075. <http://dx.doi.org/10.1111/j.1551-6709.2010.01167.x>
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*, 1082–1089. <http://dx.doi.org/10.1111/j.1467-9280.2006.01834.x>

- Davis, J. I., Gross, J. J., & Ochsner, K. N. (2011). Psychological distance and emotional experience: What you see is what you get. *Emotion, 11*, 438–444. <http://dx.doi.org/10.1037/a0021783>
- DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behavior, 32*, 204–215. <http://dx.doi.org/10.1016/j.evolhumbehav.2011.01.003>
- DeScioli, P., Gilbert, S. S., & Kurzban, R. (2012). Indelible victims and persistent punishers in moral cognition. *Psychological Inquiry, 23*, 143–149. <http://dx.doi.org/10.1080/1047840X.2012.666199>
- Deska, J. C., Kunstman, J. W., Bernstein, M. J., Ogungbadero, T., & Hugenberg, K. (2020). Black racial phenotypicity shapes social pain and support judgments. *Journal of Experimental Social Psychology, 90*, 103998. <http://dx.doi.org/10.1016/j.jesp.2020.103998>
- Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology, 78*, 1135–1149. <http://dx.doi.org/10.1037/0022-3514.78.6.1135>
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*, 107–143. <http://dx.doi.org/10.1111/j.1756-8765.2008.01006.x>
- Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75*, 617–638. <http://dx.doi.org/10.1037/0022-3514.75.3.617>
- Gilovich, T., & Medvec, V. H. (1995). The experience of regret: What, when, and why. *Psychological Review, 102*, 379–395.
- Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology, 78*, 211–222. <http://dx.doi.org/10.1037/0022-3514.78.2.211>
- Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75*, 332–346. <http://dx.doi.org/10.1037/0022-3514.75.2.332>
- Gray, K. (2012). The power of good intentions: Perceived benevolence soothes pain, increases pleasure, and improves taste. *Social Psychological and Personality Science, 3*, 639–645. <http://dx.doi.org/10.1177/1948550611433470>
- Gray, K., & Schein, C. (2012). Two minds vs. two philosophies: Mind perception defines morality and dissolves the debate between deontology and utilitarianism. *Review of Philosophy and Psychology, 3*, 405–423. <http://dx.doi.org/10.1007/s13164-012-0112-5>
- Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry, 23*, 206–215. <http://dx.doi.org/10.1080/1047840X.2012.686247>
- Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science, 19*, 1260–1262. <http://dx.doi.org/10.1111/j.1467-9280.2008.02208.x>
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*, 101–124. <http://dx.doi.org/10.1080/1047840X.2012.651387>
- Green, J. D., Davis, J. L., Luchies, L. B., Coy, A. E., Van Tongeren, D. R., Reid, C. A., & Finkel, E. J. (2013). Victims versus perpetrators: Affective and empathic forecasting regarding transgressions in romantic relationships. *Journal of Experimental Social Psychology, 49*, 329–333. <http://dx.doi.org/10.1016/j.jesp.2012.12.004>
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*, 364–371. <http://dx.doi.org/10.1016/j.cognition.2009.02.001>
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400. <http://dx.doi.org/10.1016/j.neuron.2004.09.027>
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108. <http://dx.doi.org/10.1126/science.1062872>
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*, 382–385. <http://dx.doi.org/10.1038/nm827>
- Haggard, P., & Tsakiris, M. (2009). The experience of agency: Feelings, judgments, and responsibility. *Current Directions in Psychological Science, 18*, 242–246. <http://dx.doi.org/10.1111/j.1467-8721.2009.01644.x>
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.
- Heider, F. (1958). *The psychology of interpersonal relations*. Hoboken, NJ: Wiley <http://dx.doi.org/10.1037/10628-000>
- Helzer, E. G., & Dunning, D. (2012). Why and when peer prediction is superior to self-prediction: The weight given to future aspiration versus past achievement. *Journal of Personality and Social Psychology, 103*, 38–53. <http://dx.doi.org/10.1037/a0028124>
- Hindriks, F. (2008). Intentional action and the praise-blame asymmetry. *The Philosophical Quarterly, 58*, 630–641. <http://dx.doi.org/10.1111/j.1467-9213.2007.551.x>
- Hindriks, F., Douven, I., & Singmann, H. (2016). A new angle on the Knobe effect: Intentionality correlates with blame, not with praise. *Mind & Language, 31*, 204–220. <http://dx.doi.org/10.1111/mila.12101>
- Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between Blacks and Whites. *Proceedings of the National Academy of Sciences of the United States of America, 113*, 4296–4301. <http://dx.doi.org/10.1073/pnas.1516047113>
- Hoffman, M., Yoelli, E., & Navarrete, C. (2016). Game theory and morality. In T. K. Shackelford & R. D. Hansen (Eds.), *The evolution of morality* (pp. 289–316). New York, NY: Springer. http://dx.doi.org/10.1007/978-3-319-19671-8_14
- Hume, D. (1740). *A treatise of human nature*. New York, NY: Dover.
- Igou, E. R. (2008). “How long will I suffer?” versus “How long will you suffer?” A self-other effect in affective forecasting. *Journal of Personality and Social Psychology, 95*, 899–917. <http://dx.doi.org/10.1037/a0011619>
- Kahneman, D., & Snell, J. (1990). Predicting utility. In R. Hogarth (Ed.), *Insights in decision making* (pp. 295–310). Chicago, IL: University of Chicago Press.
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*, 190–194. <http://dx.doi.org/10.1093/analysis/63.3.190>
- Kross, E., & Ayduk, O. (2008). Facilitating adaptive emotional analysis: Distinguishing distanced-analysis of depressive experiences from immersed-analysis and distraction. *Personality and Social Psychology Bulletin, 34*, 924–938. <http://dx.doi.org/10.1177/0146167208315938>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134. <http://dx.doi.org/10.1037/0022-3514.77.6.1121>
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International Affective Picture System (IAPS): Technical manual and affective ratings*. Gainesville, FL: NIMH Center for the Study of Emotion and Attention, University of Florida.
- Lau, T., Morewedge, C. K., & Cikara, M. (2016). Overcorrection for social-categorization information moderates impact bias in affective forecasting. *Psychological Science, 27*, 1340–1351. <http://dx.doi.org/10.1177/0956797616660292>
- Leitner, J. B., Ayduk, O., Mendoza-Denton, R., Magerman, A., Amey, R., Kross, E., & Forbes, C. E. (2017). Self-distancing improves interpersonal perceptions and behavior by decreasing medial prefrontal cortex

- activity during the provision of criticism. *Social Cognitive and Affective Neuroscience*, 12, 534–543. <http://dx.doi.org/10.1093/scan/nsw168>
- Leyens, J. P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to in-groups and out-groups. *Personality and Social Psychology Review*, 4, 186–197. http://dx.doi.org/10.1207/S15327957PSPR0402_06
- Loewenstein, G. F., & Schkade, D. (1999). Wouldn't it be nice? Predicting future feelings. In D. Kahneman, E. Diener, & N. Schwartz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 85–105). New York, NY: Russell Sage Foundation.
- Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin*, 132, 895–919. <http://dx.doi.org/10.1037/0033-2909.132.6.895>
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101–121. <http://dx.doi.org/10.1006/jesp.1996.1314>
- McRae, K., Ciesielski, B., & Gross, J. J. (2012). Unpacking cognitive reappraisal: Goals, tactics, and outcomes. *Emotion*, 12, 250–255. <http://dx.doi.org/10.1037/a0026351>
- Miller, R., & Cushman, F. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social and Personality Psychology Compass*, 7, 707–718. <http://dx.doi.org/10.1111/spc3.12066>
- Moons, W. G., Chen, J. M., & Mackie, D. M. (2017). Stereotypes: A source of bias in affective and empathic forecasting. *Group Processes & Intergroup Relations*, 20, 139–152. <http://dx.doi.org/10.1177/1368430215603460>
- Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, 84, 221–236. [http://dx.doi.org/10.1016/S0010-0277\(02\)00048-3](http://dx.doi.org/10.1016/S0010-0277(02)00048-3)
- Pfister, R., Obhi, S. S., Rieger, M., & Wenke, D. (2014). Action and perception in social contexts: Intentional binding for social action effects. *Frontiers in Human Neuroscience*, 8, 667. <http://dx.doi.org/10.3389/fnhum.2014.00667>
- Pollmann, M. M. H., & Finkenauer, C. (2009). Empathic forecasting: How do we predict other people's feelings? *Cognition and Emotion*, 23, 978–1001. <http://dx.doi.org/10.1080/02699930802264895>
- Preston, J., & Wegner, D. M. (2007). The eureka error: Inadvertent plagiarism by misattributions of effort. *Journal of Personality and Social Psychology*, 92, 575–584. <http://dx.doi.org/10.1037/0022-3514.92.4.575>
- Reit, E. S., & Critcher, C. R. (2020). The commonness fallacy: Commonly chosen options have less choice appeal than people think. *Journal of Personality and Social Psychology*, 118, 1–21. <http://dx.doi.org/10.1037/pspa0000172>
- Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, 80, 340–352. <http://dx.doi.org/10.1037/0022-3514.80.2.340>
- Schein, C., Goranson, A., & Gray, K. (2015). The uncensored truth about morality. *The Psychologist*, 28, 982–985.
- Schein, C., & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Personality and Social Psychology Bulletin*, 41, 1147–1163. <http://dx.doi.org/10.1177/0146167215591501>
- Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, 22, 32–70. <http://dx.doi.org/10.1177/1088868317698288>
- Schein, C., Hester, N., & Gray, K. (2016). The visual guide to morality: Vision as an integrative analogy for moral experience, variability and mechanism. *Social and Personality Psychology Compass*, 10, 231–251. <http://dx.doi.org/10.1111/spc3.12247>
- Shultz, T. R. (1980). Development of the concept of intention. In W. A. Collins (Ed.), *The Minnesota Symposium on child psychology* (Vol. 13, pp. 131–158). Hillsdale, NJ: Erlbaum.
- Shultz, T. R., & Wright, K. (1985). Concepts of negligence and intention in the assignment of moral responsibility. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 17, 97–108.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). *Life after p-hacking* [PowerPoint slides]. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205186
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27, 76–105. [http://dx.doi.org/10.1016/0022-1031\(91\)90011-T](http://dx.doi.org/10.1016/0022-1031(91)90011-T)
- Steffel, M., Williams, E. F., & Permann-Graham, J. (2016). Passing the buck: Delegating choices to others to avoid responsibility and blame. *Organizational Behavior and Human Decision Processes*, 135, 32–44. <http://dx.doi.org/10.1016/j.obhdp.2016.04.006>
- Steinmetz, J., Touré-Tillery, M., & Fishbach, A. (2020). The first-member heuristic: Group members labeled “first” influence judgment and treatment of groups. *Journal of Personality and Social Psychology*, 118, 706–719. <http://dx.doi.org/10.1037/pspi0000201>
- Thompson, S. C. (1999). Illusions of control: How we overestimate our personal influence. *Current Directions in Psychological Science*, 8, 187–190. <http://dx.doi.org/10.1111/1467-8721.00044>
- Ubel, P. A., Loewenstein, G., Schwarz, N., & Smith, D. (2005). Misimagining the unimaginable: The disability paradox and health care decision making. *Health Psychology*, 24, S57–S62. <http://dx.doi.org/10.1037/0278-6133.24.4.S57>
- Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, 29, 1159–1168. <http://dx.doi.org/10.1177/0146167203254597>
- van Dijk, W. W., Van Dillen, L. F., Seip, E. C., & Rotteveel, M. (2012). Emotional time travel: Emotion regulation and the overestimation of future anger and sadness. *European Journal of Social Psychology*, 42, 308–313. <http://dx.doi.org/10.1002/ejsp.1853>
- Walsh, E., & Ayton, P. (2009). What would it be like for me and for you? Judged impact of chronic health conditions on happiness. *Medical Decision Making*, 29, 15–22. <http://dx.doi.org/10.1177/0272989X08326147>
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press. <http://dx.doi.org/10.7551/mitpress/3650.001.0001>
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92, 548–573.
- Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, 14, 131–134. <http://dx.doi.org/10.1111/j.0963-7214.2005.00355.x>
- Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, 78, 821–836. <http://dx.doi.org/10.1037/0022-3514.78.5.821>
- Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, 100, 283–301. <http://dx.doi.org/10.1016/j.cognition.2005.05.002>
- Yamagishi, T., Terai, S., Kiyonari, T., Mifune, N., & Kanazawa, S. (2007). The social exchange heuristic: Managing errors in social exchange. *Rationality and Society*, 19, 259–291. <http://dx.doi.org/10.1177/1043463107080449>

Received September 18, 2018

Revision received July 5, 2020

Accepted July 30, 2020 ■