Collective Responsibility: The Individual Actor’s Perspective

Peter Kardos\textsuperscript{1}
Bernhard Leidner\textsuperscript{1}
Emanuele Castano\textsuperscript{2}
Brian Lickel\textsuperscript{1}

1: University of Massachusetts Amherst
2: New School for Social Research

Correspondence:
Peter Kardos
Department of Psychology
University of Massachusetts Amherst
135 Hicks Way, Tobin Hall, Amherst, MA 01003
pkardos@psych.umass.edu
Abstract

Collective responsibility processes have been investigated from the perspectives of the outgroup and the ingroup. When an individual wrongdoers’ group membership is salient, outgroup members often generalize the wrongdoer’s behavior to the wrongdoer’s entire group and blame it collectively. Likewise, fellow ingroup members may feel tarnished by the action and punish the wrongdoer who brought blame and disrepute to the group. This paper extends theory and research on collective responsibility with a third perspective, namely that of the individual actor whose behavior triggers the attribution of collective blame. Five experiments tested the hypotheses that collective responsibility processes influence the individual actors’ appraisals, behavior and emotions. The possibility of collective blame for their individual action prompted more prosocial behavior among participants (Experiment 1 and 2). Participants also experienced more ingroup reputation concern and in turn more negative emotions (Experiment 3-5) for a past wrongdoing if it could reflect negatively on the ingroup in the eyes of outgroups. The increased negative emotions then motivated participants to improve the ingroup’s image (Experiment 5). The effects were further moderated by perceived ingroup entitativity (Experiment 4).

Keywords: collective responsibility; collective blame; intergroup context; reputation; entitativity
Individual wrongdoing often has group level implications. When a wrongdoer’s membership in a social group is salient, people often make inferences about individual transgressions that are linked to the wrongdoer’s group memberships. Whether they were harmed or not, outsiders may not only judge the individual actor for his or her actions, but also judge and blame the individual’s entire group (Hamill, Wilson, & Nisbett, 1980; Lickel, Schmader, & Hamilton, 2003), under some circumstances to the point of targeting the entire group for retaliation or censure (Lickel, 2012). Members of the wrongdoer’s group might sometimes support the wrongdoer’s actions, if they were committed on behalf of the group. But extensive research shows the limits of this ingroup support. In many instances ingroup members react with feelings of guilt, shame, and anger because of the ingroup wrongdoing (Lickel, Schmader, & Spanovic, 2007), and may try to protect the ingroup’s reputation by sanctioning or excluding the wrongdoer from the group (Castano, Paladino, Coull, & Yzerbyt, 2002; Marques, Yzerbyt, & Leyens, 1988). To date, the mechanisms underlying collective responsibility have been investigated extensively from the perspective of the outgroup (i.e. collective blame) and the ingroup (i.e. group based guilt and shame, black sheep effect). They have not been investigated, however, from the perspective of the individual actor whose behavior may or may not start the collective responsibility process, triggering ingroup and/or outgroup reactions.

We extend theory and research on collective responsibility by investigating this third perspective of the collective responsibility process. We argue that people are generally aware that their individual actions may lead to collective blame. This awareness should thus influence the individual actor’s behavior. Specifically, we propose that in contexts in which group memberships are salient and the ingroup might
therefore be blamed for an individual actor’s wrongdoing by an outgroup, the individual actor will likely appraise – and possibly modify – his or her anticipated behavior in terms of its (potential) effect on the ingroup’s reputation. The recognition of intergroup situations and one’s behavior’s effect on the ingroup’s reputation may of course sometimes occur only after the blameworthy event has happened, as those contextual cues might have not been apparent or shaped yet and because people might fail to control their behavior (Baumeister & Heatherton, 1996). Yet, even in these situations, the individual actor should be impacted by collective responsibility processes, experiencing negative emotions as a result of harming the ingroup’s reputation and thus being motivated to behave better in the future. In other words, the individual actor should be subject to collective responsibility processes before and after the event.

**Past research on collective blame: Outgroup and ingroup perspectives**

People tend to perceive themselves and others in terms of group membership (Fiske & Taylor, 2013). This tendency is strongest in intergroup contexts, which highlight social categories and increase the awareness of group boundaries and differences (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). When social categories are not salient (i.e. in *intragroup* contexts), people perceive themselves and others as unique individuals. When social categories are salient (i.e. in *intergroup* contexts), however, people perceive themselves and others as members of groups instead of unique individuals (Tajfel & Turner, 1986; Turner et al., 1987). In the latter case, individual behavior can bring outgroup blame and attribution of collective responsibility to the ingroup, as well as elicit guilt and shame among fellow ingroup members.

*Outgroup perspective.* From the outgroup’s perspective, collective responsibility
is commonly reflected in holding an entire group responsible for the behavior of one or a few of its members (i.e. collective blame). Any cue regarding a perpetrator’s assumed association with different groups can trigger the blaming of these groups (Doosje & Branscombe, 2003; Doosje, Zebel, Scheermeijer, & Mathyi, 2007; Lickel et al., 2003; Lickel, 2012), even leading to retaliation against group members of the perpetrator (Lickel, Miller, Stenstrom, Denson, & Schmader, 2006). More recent theorizing on collective responsibility extended collective blame stemming from causal inferences—that is, outgroup members blame the group for having (indirectly) caused its individual member’s behavior—to collective blame stemming from trait inferences—that is, outgroup members blame the group in part for having negative traits that can be assigned to the group on the basis of the group member's behavior (Lickel & Onuki, 2015).

When making trait inferences, people use information about a person’s behavior to form impressions about his or her group (Hamilton & Sherman, 1996). This phenomenon has been a consistent finding in research on person to group generalization (Crawford, Sherman, & Hamilton, 2002; Hamill, Wilson, & Nisbett, 1980; Henderson-King & Nisbett, 1996; Rothbart & Lewis, 1988). While person to group generalization is most often interpreted in terms of cognitive heuristics, Lickel and Onuki (2015) proposed that it can also be understood as an instance of collective responsibility because the outgroup often denounces the entire group in order to explain an individual member’s behavior. Therefore, forming impressions about a group based on the behavior of one of its members may not only lead to holding that group collectively responsible, but the impact on that group’s reputation itself may constitute collective blame. For example, when a Chinese student painted graffiti on the wall of the Luxor Temple while visiting
COLLECTIVE RESPONSIBILITY

Egypt, the international media reacted by condemning the barbaric demeanor of Chinese tourists in general (BBC, 2013). The collective blame that the Chinese student’s behavior brought upon China was the negative impact on its reputation. Regardless of what type of inferences collective blame is based on, when collective blaming occurs the entire (blamed) group bears the consequences of the individual behavior. Importantly, it is the salience of the individual actor’s group membership that makes this blaming possible.

*Ingroup perspective.* From the ingroup’s perspective, collective responsibility is commonly reflected in the ingroup’s reaction to the ingroup member’s behavior that might, or already did, lead to collective blame. Ingroup members—fellow group members of the individual actor—respond to actual or possible outgroup blame in several different ways. First, ingroup members often feel negative emotions when the individual group members’ actions may bring blame or disrepute to the group from outsiders (Branscombe & Doosje, 2004; Castano & Giner-Sorolla, 2006; Doosje, Branscombe, Spears, & Manstead, 1998; Giner-Sorolla, 2012; Iyer, Leach & Pedersen, 2004; Lickel, Schmader, Curtis, Scarnier, & Ames, 2005; Lickel, Steele, & Schmader, 2011; Scarnier, Schmader, & Lickel, 2009; Wohl, Branscombe, & Klar, 2006). For example, although self-criticism for the group’s past transgression tends to be rather low (Leach, Bou Zeineddine, & Čehajić-Clancy, 2013), many present-day Dutch citizens feel guilt for the atrocities that their country committed during the colonization of Indonesia (Doosje et al., 1998). Similarly, many Americans felt shame, guilt, and anger about the American military occupation of Iraq, particularly when they thought Iraqis viewed Americans’ actions as reflecting on America’s moral character (Iyer, Schmader, & Lickel, 2007).

The ingroup members thus try to avoid the ingroup getting into situations when they
would feel guilty about it (Shepherd, Spears, & Manstead, 2013a; Shepherd, Spears, & Manstead, 2013b).

Second, ingroup members might also try to protect the ingroup’s reputation by psychologically, symbolically, or factually distancing themselves from the ingroup member whose behavior is threatening the ingroup’s reputation (Castano et al., 2002; Eidelman & Biernat, 2003; Marques & Paez, 1994; Marques et al., 1988). The misbehaving ingroup member constitutes a threat to the ingroup’s image, and can thus become the target of extreme remedial actions (Branscombe, Wann, Noel, & Coleman, 1993; Coull, Yzerbyt, Castano, Paladino, & Leemans, 2001).

*Entitativity.* Different groups attract collective blame to different degrees. The likelihood of collective blame and ingroup reactions to it depend on the level of the perceived interdependence between, and underlying common characteristics of, the group members, which is referred to as the group’s *entitativity* (Campbell, 1958; Castano, Yzerbyt & Bourguignon, 2003; Hamilton, Sherman, & Lickel, 1998; Lickel, Hamilton, & Sherman, 2001; Lickel, Hamilton, Wieczorkowska, Lewis, Sherman, & Uhles, 2000; Yzerbyt, Corneille, & Estrada, 2001). Specifically, outgroup members have to perceive the wrongdoer’s group as forming a real, cohesive group to attribute collective responsibility, and ingroup members have to do the same to worry about being held collectively responsible. Indeed, perceived group entitativity contributes to greater activation of collective responsibility processes both in terms of collective blaming by the outgroup (Denson, Lickel, Curtis, Stenstrom, & Ames, 2006; Lickel, Rutchick, Hamilton & Sherman, 2006; Lickel et al., 2003; Waytz & Young, 2012) and reactions to the individual wrongdoer by the ingroup (Abrams, Marques, Randsley de Moura, Hutchinson,
& Bown, 2004; Lickel et al., 2005).

The work reviewed above testifies to the considerable attention that has been paid to attribution of blame and responsibility at the collective level, as well as people’s reaction to wrongdoings of fellow ingroup members. Much less is known about the cognitive, affective and behavioral processes of the individual group member whose behavior triggers the collective responsibility processes among outgroup and ingroup members.

**Collective blame: The individual actor’s perspective**

Given people’s propensity to engage in collective responsibility processes both as ingroup and outgroup members, we expected individual actors to be aware that their behavior can reflect on their group and trigger collective responsibility processes. In other words, people should have an intuitive understanding that they bear responsibility for how their group will be seen by outgroup members. Prior research on stereotype threat processes also imply that people are aware of the possibility that individual behavior can affect the ingroup’s reputation—at least in terms of its preexisting stereotypes (Bosson, Haymovitz, & Pinel, 2004; Cohen & Garcia, 2005; Schmader & Johns, 2003; Schmader & Lickel, 2006a; Steele, 1997). In particularly, it has been theorized that people will experience group-reputation threat, a specific type of stereotype threat, when their behavior could confirm negative stereotypes about their group in the eyes of outgroup members (Shapiro, 2012; Shapiro & Neuberg, 2007). We argue, therefore, that people’s intuitive understanding of collective responsibility processes (Lickel et al., 2001) also includes the knowledge that their own behavior can trigger these processes. However, we hypothesize that concern about negatively influencing the
ingroup’s reputation should even occur as a more general phenomenon, without an existing stigma or group stereotype.

People’s understanding of their role as an individual in triggering collective responsibility processes among outgroup and ingroup members should manifest in a specific set of appraisals, affective reactions, and behavioral responses. Furthermore, if this understanding is truly an aspect of group and intergroup psychology, then actors’ responses should be sensitive to the context (whether it is an intergroup context or not) and should also be sensitive to the nature of the ingroup (whether it is a group that is likely to be viewed by perceivers as an appropriate target for collective blame). Finally, we argue that actors’ behavior may be influenced by collective blame processes both before an action (i.e., anticipating collective blame and modifying behavior to prevent it) as well as after an action (i.e., responding adaptively when the actor realizes in hindsight that collective blame may have occurred because of his/her individual behavior.) Below, we present a set of specific hypotheses about the appraisal, emotional and behavioral responses of the individual actor when collective blame processes are activated.

**Appraisals.** As collective responsibility influences a behavior’s possible consequences, it should also change how the individual actor perceives the behavior. Particularly, collective responsibility renders a negative behavior relevant to how the ingroup is perceived by others and should thus activate thoughts and concerns in the actor about the ingroup’s reputation. People are generally sensitive to the ingroup’s reputation, most prominently in terms of the group’s morality (Branscombe, Ellemers, Spears, & Doosje, 1999; Leach, Ellemers & Barreto, 2007; Tajfel & Turner, 1979) and tend to see events in light of the likelihood to trigger evaluations of the ingroup by other groups
(Marques, Abrams, & Serôdio, 2001). Recognizing that their own behavior could reflect on the ingroup should increase their concern over the ingroup’s reputation. We hypothesized that when collective blame is possible, people are more likely to appraise their own behavior in terms of its potential negative effects on the ingroup’s reputation.

**Behavior.** Appraising their own behavior in terms of its negative group level consequences should influence behavior. It is a basic human motivation that people want to view their own group as good or better than other groups and they also want other groups to share that view (Tajfel, 1982). People care about the ingroup’s image and want to protect it. Understanding that they are perceived as representative of the ingroup and so that their own behavior may influence the ingroup’s reputation should prompt people to behave better: either by refraining from negative behavior or by executing positive behavior. In this respect, the possibility of collective blame may actually promote prosocial behavior. We thus hypothesized that people behave more positively when collective blame is possible.

**Emotions.** Despite their motivation to protect the group, people may nonetheless engage in actions that can tarnish the group’s image either because of insufficient self-control or because they realize too late that the behavior could have activated collective blame. In this case, people should (re)appraise their past behavior in terms of collective responsibility processes and become concerned about having damaged the ingroup’s reputation. They may experience negative emotions for the harmful consequences of the behavior regardless of its impact on the ingroup image, but the possibility of activating outgroup blame and thereby damaging the ingroup image should amplify this experience. The negative emotions should then have further consequences. In order to avoid
experiencing them, people should be motivated to behave more positively in subsequent situations that bear similar consequences. Therefore, we hypothesized that people would feel more negative emotions about a past behavior if it could activate collective blame, and that the increased negative emotions, in turn, would motivate more positive subsequent behavior.

**Overview of the present research**

The hypotheses identified above were tested across five experiments in which we manipulated the likelihood that a person’s individual behavior could activate collective blame for the ingroup. We operationalized collective blame possibility by manipulating if the outgroup could make trait inferences about the individual actor’s group. Each experiment was designed to manipulate only the group reputational consequences of one’s behavior, while the possible personal consequences of the behavior were kept constant across conditions.

In Experiment 1 participants imagined being in a decision making situation where, depending on condition, their engaging in a negative behavior could be witnessed by an outgroup vs. ingroup member. Experiment 2 combined a minimal group paradigm with a behavioral interaction paradigm. Participants played a dictator game and we assessed how prosocially they behaved depending on the possibility that their behavior could influence the ingroup’s image, and how they felt about their selfish or generous behavior depending on its likelihood of triggering collective blame. Experiment 3 had participants imagine committing a transgression in a foreign country, manipulating whether or not their group membership became known to outgroup members, and measured ingroup reputation concern and emotional reactions. Using a similar imagined scenario,
Experiment 4 manipulated whether ingroup members or outgroup members witnessed the wrongdoing and measured ingroup reputation concern and emotional reactions, while also testing the moderating role of perceived ingroup entitativity. Experiment 5 used a behavioral interaction paradigm manipulating whether or not participants represented the ingroup before an outgroup, and measuring ingroup reputation concern, emotional reactions as well as the emotions’ consequences for subsequent behavior.

Statistical power. We performed post-hoc power analyses using the G*Power program (Faul, Erdfelder, Lang, & Buchner, 2007). Based on the obtained effect and sample sizes, the power to detect the effects reported below was .73 for Experiment 1, .78 for Experiment 2, .78 again for Experiment 3, .53 for Experiment 4, and .79 for Experiment 5. These power indices are greater than the average power of .35 in social psychological studies (Bakker, van Dijk, & Wicherts, 2012; Marszalek, Barber, Kohlhart, & Holmes, 2011), and also on average above the average power of .65 in studies published in upper tier social psychology journals including Journal of Personality and Social Psychology (Fraley & Vazire, 2014).

Experiment 1

Experiment 1 tested the extent to which people are sensitive to the possibility of collective blame and whether people would be more reluctant to commit a negative behavior in contexts where collective blame is possible (intergroup) than in contexts where it is not possible (intragroup). To create intergroup vs. intragroup contexts we manipulated whether an individual transgression could be judged only by members of the ingroup or also by a member of the outgroup. We hypothesized that people would intend to behave more prosocially when the outgroup member could form a negative opinion
about their ingroup.

**Method**

**Participants.** Of 100 Americans recruited via Amazon Mechanical Turk, four incorrectly remembered the manipulation material in the manipulation test and were dropped, leaving 96 participants for the analyses ($M_{age} = 35.78$, $SD = 11.44$, range 19-66; 46 female).

**Materials and Procedure.** Following a customary method of studying people’s reactions to their own wrongdoing, we used imagined scenarios (Baumeister, Stillwell, & Heatherton, 1995; Castano & Giner-Sorolla, 2006; McGraw, 1987) in which participants were asked to read a hypothetical situation and imagine themselves in the role of the protagonist. In both conditions, participants imagined going to a farmers market and realizing there that they had left their money at home while standing hungrily before a fruit vendor. Participants were then prompted to contemplate whether they would take and eat a few grapes, risking the vendor’s outrage in case they were caught. In the intragroup condition the vendor’s nationality was not indicated, so that participants likely assumed the vendor to be American (since the farmers market was in the U.S.). In the intergroup condition the vendor was French. We hypothesized that people would be less willing to steal a few grapes in the intergroup condition (i.e. when their behavior may reflect poorly on their ingroup in the eyes of an outgroup member).

**Behavioral intention.** Participants were asked “How likely is it that you would have decided to take the grapes?” The answers were measured on an 8-point continuous scale from 1 = Very unlikely to 8 = Very likely ($M = 1.71$, $SD = 1.10$).
**Manipulation check.** The manipulation check asked participants to recall the vendor’s nationality and select the correct answer from “American”, “French” and “Do not know”.

**Results**

**Behavioral intention.** The manipulation significantly influenced participants’ intentions to take the grapes, $F(1, 94) = 4.11, p = .045, \eta^2 = .04$. Despite an apparent floor effect – participants being reluctant to take the grapes in both conditions – their intentions to take the grapes were significantly lower in the intergroup ($M = 1.47, SD = .80$) than in the intragroup condition ($M = 1.92, SD = 1.28$).

**Discussion**

The result that people intended to behave in a less negative way in the eyes of an outgroup rather than ingroup member provided initial support for our hypothesis that collective responsibility promotes prosocial behavior.

**Experiment 2**

Experiment 1 showed that when people imagined themselves being tempted to commit a wrongdoing, their imagined behavior was more restrained in an intergroup compared to an intragroup context. In Experiment 2, we developed a modified dictator game (Camerer, 2003; Forsythe, Horowitz, Savin, & Sefton, 1994) to test the effect of potential collective blame on actors’ *actual* behavior. The dictator game is believed to assess norms of generosity and to index altruistic pro-social behavior that is not contingent on concern about punishment or explicit desire for reciprocity. In our paradigm, we introduced a reputation system (cf. Feinberg, Willer, & Schultz, 2014) and manipulated whether individual selfishness could or could not influence the reputation of
participants’ ingroup in the eyes of outgroup members. Furthermore, to show that actors’ behavioral sensitivity to the possibility of collective blame does not depend on extensive intergroup history or group stereotypes, we used a minimal group paradigm (Tajfel, 1970; Tajfel, Billig, Bundy, & Flament, 1971) where participants were assigned to play as members of a blue or a red team. Besides assessing participants’ actual behavior, we also assessed their emotional response, predicting that people who behaved more selfish would feel worse—but only when their behavior could damage the ingroup’s reputation.

Method

Participants. Seventy-eight college undergraduates participated for research credit as well as for a $2 monetary reward. Four participants who raised suspicions whether our cover story was true were dropped, leaving seventy-four participants for subsequent analyses ($M_{age} = 19.89, SD = 1.44$, range 18-25 years; 60 female).

Materials and procedure. According to the cover story, participants played against each other as members of opposing teams (blue team vs. red team). In reality, there were no other players. Participants were first asked to draw a card from a bag to determine which team they were going to join and in which round of the game they would play. Unbeknownst to participants, all cards in the bag said “Blue 1,” assigning them to the blue team and the first game to be played. Participants were asked to keep the card during the entire study and to put a sticky note with “Blue 1” written on it on their shirt. These steps served to establish participants’ sense of membership in the blue group. Participants were told that there were four blue players in the blue group and four red players in the red group; that in each game round, one blue player would play against one red player; and that as “Blue 1,” the participant was to be the first blue player to play
The experimenter then led the participant through a hallway with doors with signs from “Blue 1” to “Blue 4” and entered the room with the sign “Blue 1.” Participants were seated before a computer and told to follow the instructions in the computer-assisted survey. To further bolster our cover story, the first two questions in the survey (“Which group does the participant belong to?” and “What number in the Blue group does the participant have?”) were answered by the experimenter.

Playing the dictator game. Participants were given detailed instructions as to how to play the game. Participants, as “Blue 1,” were told to play with “Red 1.” Participants were told that they were the allocator and “Red 1” the receiver. As allocator, they got 20 aluminum coins, each worth 10 cents. They had to anonymously divide the coins between themselves and “Red 1” as they wished. Participants were told that “Red 1” could not respond or contact them, that they would lose whatever they chose to give to “Red 1,” and that the coins they owned at the end of the game would be exchanged for real money. To ensure anonymity, a coin acceptor machine was connected to the computer and participants had to throw in the coins that they wanted to give to “Red 1.” The coin acceptor would count the coins and send a message to “Red 1,” ostensibly sitting in another room, informing “Red 1” about how many coins he/she received. The coins that participants wanted to keep for themselves had to be placed in an envelope, to be exchanged for real money by a different experimenter later on. Participants were further told that after they would finish their game with “Red 1,” “Blue 2” would play with “Red 2,” but switching roles so that “Red 2” would be the allocator and “Blue 2” the receiver. Then “Blue 3” would play with “Red 3,” switching roles again, and so on for the fourth pair.
Manipulating collective responsibility in the dictator game. To manipulate participants’ perception of whether their individual behavior could or could not influence the ingroup’s (i.e. blue team’s) reputation, we built a reputation system within the game. Besides of the game instructions we detailed above, in the collective responsibility condition participants were additionally told that at the end of each turn (e.g. after “Blue 1” played “Red 1”), the receivers (e.g. “Red 1”) could send a message to their own group members about the allocator’s behavior, where the message indicated the allocator’s group membership (i.e. “Blue 1” was selfish or generous). In the control condition there was no such opportunity for the receivers.

After reading the rules, participants’ understanding of them was tested with ten questions. After incorrect answers, the survey showed the correct answer and explained again the specific detail. Participants then notified the experimenter, who gave them the 20 coins. The experimenter left participants alone to allocate the coins between themselves and “Red 1.” Participants then continued the survey, completing a measure of state emotions.

Emotions. Eight items measured how participants felt after the game: six negative emotions (guilt, remorse, shame, sorry, worry, anxiety), and two positive emotions (satisfaction, happiness) for control purposes and because participants could feel good about earning money. All emotions were rated on continuous scales from 1 (Not at all) to 9 (Very much). Factor analyses suggested two meaningful factors. One of them comprised the six negative emotions ($M = 2.02$, $SD = 1.31$, $\alpha = .89$), the other the two positive emotions ($M = 5.65$, $SD = 1.72$, $\alpha = .87$).

At the end, regardless of their behavior in the game, participants received the
maximum possible payout of $2. After participants were debriefed and left, the experimenter opened the envelopes and recorded how many coins participants had kept for themselves, which formed our (behavioral) dependent variable.

**Results**

Participants kept significantly fewer coins for themselves in the collective responsibility ($M = 10.39$) than in the control condition ($M = 12.08$), $F(1, 72) = 7.53, p = .007, \eta^2 = .09$, indicating that they behaved more generously when their decision could influence the ingroup’s reputation.

Next we tested how emotions were influenced by one’s selfish behavior depending on the behavior’s possible group level consequences. If people thought that they had behaved selfishly (rather than generously), they should have felt worse, and especially so when this selfish behavior could influence the ingroup’s reputation (rather than not). In other words, we predicted that the effect of selfish behavior would depend on the presence vs. absence of collective responsibility. To test this hypothesis, we ran a moderated regression with condition as categorical and selfish behavior (i.e. number of coins participants had kept for themselves [centered]) as continuous IVs (and their interaction), and negative emotions as DV. Selfish behavior had a significant main effect on negative emotions, $F(1, 70) = 10.13, p = .002, \eta^2 = .13$; the more selfish participants’ behavior had been, the more negative emotions they felt ($\beta = .54$). The main effect of collective responsibility did not reach significance, but showed a trend, $F(1, 70) = 2.68, p = .106, \eta^2 = .04$, with participants reporting more negative emotions when their behavior could reflect on the ingroup ($M = 1.89$) rather than when it could not ($M = 1.67$). Most importantly, these main effects were qualified by the hypothesized two-way interaction
between collective responsibility and selfish behavior, $F(1, 70) = 11.64, p = .001, \eta^2 = .14$. Participants who behaved relatively generously reported similar levels of negative emotions when their behavior could reflect on the ingroup ($M = 2.01$) and when it could not ($M = 1.34$), $t(70) = 1.57, p = .12$. Participants who behaved relatively selfishly reported significantly more negative emotions when their behavior could reflect on the ingroup ($M = 3.57$) rather than when it could not ($M = 1.92$), $t(70) = 3.40, p = .001, d = .6$ (see Figure 1). Importantly, this interaction was driven by the combination of collective responsibility and selfish behavior.

**Discussion**

In Experiment 2 we developed a behavioral paradigm for testing how collective responsibility influences selfish versus generous behavior. As hypothesized, people were more generous to an anonymous outgroup person when their behavior could reflect on their ingroup. In doing so, participants acted against their self-interest (earning money) to protect the ingroup’s image. Thus, even in minimal groups, concerns about the ingroup’s reputation can induce changes in behavior even when those changes incur real costs. Also, not only was selfish behavior lower in the collective responsibility condition, but players who “failed” to heed the potential consequences of their actions for the group and behaved selfishly felt worse in that condition than in the control condition. These two findings highlight that actors’ concern for the ingroup’s reputation plays a role both before action (often modifying behavior to protect the group image) and after action (showing negative feelings and self-sanctions when having committed a blameworthy action that reflects on the group). Experiments 3-5 further unpacked the appraisals, emotions, and future behavior that occur when actors fail to protect the ingroup from
collective blame.

**Experiment 3**

In Experiments 1 and 2 participants behaved more positively when their behavior could affect the ingroup’s reputation. But, people are not always fully able to control their behavior. Furthermore, people often realize too late that their actions might have had group level consequences. Collective responsibility processes should nonetheless influence the individual actors’ responses to their transgressions. Experiment 3 tested how people think and feel learning that their past behavior could trigger collective blame. We hypothesized that when their behavior could trigger collective blame, people would be more concerned about the ingroup’s reputation and, in turn, feel more negative emotions. While Experiment 2 already gave preliminary evidence for this hypothesis, it only did so for people who spontaneously had engaged in selfish behavior that could damage the ingroup’s reputation. To rule out any possible confounds, Experiment 3 instead manipulated participants’ past behavior directly in order to obtain causal evidence.

**Method**

**Participants.** 118 Americans were recruited and compensated via Amazon Mechanical Turk ($M_{age} = 35.86$, $SD = 13.53$, range 18-68; 72 female).

**Procedure.** Participants read a short scenario, imagining they were the main character: a hiker. They imagined that during a hiking trip in China they accidentally caused fire damage to a grove of trees and a sanctuary, but got out safely and went home without anybody knowing about their role in the fire. Chinese media reported the fire and the considerable damage it caused. In the control condition, participants then read
that Chinese authorities had not been able to gather any information about the person causing the fire, but believed it was a hiker. In the experimental condition, the Chinese authorities believed that the suspect was an American hiker. Participants thus damaged the forest in both conditions but in the experimental condition their revealed group membership could additionally damage the ingroup’s reputation. All items were measured on continuous scales ranging from 0 (Not at all) to 7 (Very much), unless noted otherwise.

**Emotions.** Participants completed a self-report measure of state emotionality, rating ten emotions on scales from 0 (Not at all) to 10 (Very intensively). Six items measured negative self-conscious emotions (guilt, shame, embarrassment, disgrace, remorse, sorry) two items measured anxiety-related emotions (worry, anxiety) and two items measured a positive (nostalgia) and a negative (envy) emotion for control purposes. Factor analysis led to a three-factor solution. The six negative self-conscious emotions loaded on one factor ($\alpha = .90, M = 8.87, SD = 1.32$), the two anxiety-related emotions on a second factor ($\alpha = .96, M = 8.94, SD = 1.54$), and the two unrelated emotions on a third ($\alpha = .65, M = 1.01, SD = 1.68$). Three univariate outliers on negative self-conscious emotions (2½ standard deviations below the mean) were eliminated from further analyses (Tabachnik & Fidell, 2007).¹

**Ingroup reputation concern.** Participants appraised the consequences of their behavior answering two questions about the potential effects of the described events on the reputation of the U.S. (“Do you think that the incident described in the scenario may reflect poorly on Americans?” and “Do you think that the incident described in the scenario may somehow damage the image of the United States?”; $\alpha = .87, M = 4.89, SD$
Appraised cause of emotions. We assessed the appraised cause of the emotions to test to what extent participants attributed their negative emotions to their behavior’s group level consequences. Participants were asked why they thought that they would experience the reported emotions and then rated their level of agreement with two statements: “I would feel bad because my behavior led to the damage in the national park” ($M = 6.48, SD = 0.74$) and “I would feel bad because my behavior may reflect poorly on other Americans” ($M = 4.50, SD = 2.21$).

Attitudes toward China. To control for pre-existing attitudes toward the outgroup, three items assessed participants’ attitudes toward China in general, in cultural and in economic terms ($\alpha = .71$, $M = 3.99$, $SD = 1.18$).

Results

Emotions. A general linear model revealed a significant effect of condition on negative self-conscious emotions, $F(1, 113) = 6.72, p = .011$, $\eta^2 = .056$, but not on the anxiety-related or the unrelated emotions, $ps > .10$. People reported significantly more negative self-conscious emotions when they believed that the outgroup was aware of their group membership ($M = 9.16, SD = 0.97$) than when it was not ($M = 8.62, SD = 1.21$).

Ingroup reputation concern. Condition significantly influenced ingroup reputation concern, $F(1, 116) = 6.58, p = .012$, $\eta^2 = .054$. Participants’ belief that their actions could hurt the ingroup’s image were stronger in the experimental ($M = 5.27, SD = 1.48$) than in the control condition ($M = 4.54, SD = 1.62$).

Appraised cause of emotions. We ran a repeated-measures analysis of variance (ANOVA) with condition as independent variable (IV), the two appraised causes
(damaging the forest and damaging U.S. reputation) as dependent variables (DV$s), and the type of cause as repeated factor. As expected, the condition by type interaction was significant, $F(1, 116) = 5.71, p = .018$. Participants’ attributions of their negative self-conscious emotions to damaging the forest did not significantly differ between conditions, $F(1, 116) = 0.97, p = .326$, whereas their attributions of their negative self-conscious emotions to harming U.S. reputation were significantly stronger in the experimental ($M = 5.09$) than in the control condition ($M = 3.96$), $F(1, 116) = 8.08, p = .005, \eta^2 = .065$. This finding indicated that the condition effect on negative self-conscious emotions was carried by the concern over damage to the ingroup’s reputation, not by the concern over the damage to the outgroup or its resources (i.e. the Chinese forest).

**Mediation analysis.** In emotion theory, both on individual and on group level, the appraisal of the event is conceptualized as the trigger of the emotion (Frijda, Kuipers, & ter Schure, 1989; Kuppens & Yzerbyt, 2012; Smith, 1993). We thus tested whether concern over the ingroup’s reputation mediated the effect of condition on negative self-conscious emotions, using 95% bias-corrected confidence intervals and 5000 bootstrap resamples (Hayes, 2012, model 4). In this model, condition significantly predicted ingroup reputation concern, $t = 2.63, p < 0.01$, which in turn predicted negative self-conscious emotions, $t = 2.45, p = .016$. The confidence interval of the indirect effect did not include zero (CI$_{95} = [0.021, 0.318]$), indicating a significant indirect effect (see Figure 2).

All reported analyses held when controlling for attitudes toward China.

**Discussion**
In line with our hypothesis, imagining oneself as the perpetrator of a negative action led to more negative emotions when the outgroup learned about the perpetrator’s group membership. In both conditions participants felt negative emotions for a harmful behavior, but this response was even stronger when the behavior could also damage the ingroup’s reputation. Furthermore, as predicted, the perpetrator’s concern over his or her ingroup’s reputation mediated the effect of collective blame possibility on negative emotions. Importantly, the results lend direct support to the hypothesis that collective responsibility processes affect the individual actor not only before an action, but also after it has already happened.

**Experiment 4**

Past research on collective responsibility has found that people more likely blame high than low entitative groups (Denson, et al., 2006; Lickel, et al. 2003; Lickel, et al. 2006). The reason is that one has to assume a certain level of similarity and connection among the members of a group to judge the whole group based on one of its members’ behavior. Without these qualities, people are unlikely to make causal or trait inferences from one member’s behavior to the whole group. Accordingly, if the wrongdoer’s ingroup is less entitative, its members are less concerned about drawing collective blame (Lickel, Schmader, & Barquissau, 2004). We expect that the same logic would apply to the individual actor as well. A person should not fear that her individual behavior can tarnish the ingroup’s image if she doubts that outsiders would assume strong association between the members of the ingroup. But if the ingroup is perceived as highly entitative, then the wrongdoer should more likely expect that outsiders would generalize her behavior to her ingroup and this possibility of collective blame should increase the
reputation concern and negative emotions that the wrongdoer experiences over her transgression. In Experiment 4, we tested this hypothesis.

Experiment 4 also aimed to improve our manipulation of collective responsibility. In Experiment 3, the wrongdoing happened abroad in both conditions, possibly priming intergroup context. While this would, if anything, have worked against the effect we predicted (and indeed found), Experiment 4 avoided this possible confound by manipulating whether the same transgression happened at home or abroad, thus establishing better experimental control over whether the context was intergroup or intragroup. Also, we measured the wrongdoer’s concern about both the ingroup’s and his or her personal image in order to cleanly separate those constructs. Additionally, we assessed the perceived damage to the ingroup’s image with an alternative, less blatant measure.

Method

Participants. Of 208 Americans recruited and paid via Amazon Mechanical Turk, eighteen could not correctly recall information from the manipulation material during the manipulation check and were thus dropped, leaving 190 participants for subsequent analyses ($M_{age} = 34.96, SD = 13.43$, range 18-73; 108 female).

Materials and Procedure. Participants were randomly assigned to one of two conditions. In both conditions participants imagined being the main character of the events described in a scenario. The scenario described two friends getting into a heated argument on the street, observed by local bystanders. In one condition it happened during a vacation in a Southern Californian city (i.e. intragroup condition), whereas in the other condition it happened in an unspecified foreign country (i.e. intergroup condition).
Therefore, the bystanders witnessing the quarrel were strangers in both conditions, but it was only the intergroup condition where the witnesses were members of a national outgroup who could form a negative opinion about participants’ national ingroup.

*Emotions.* After reading the scenario, participants completed an emotion response scale, rating nine emotions on continuous scales from 0 (Not at all) to 10 (Extremely much). As in Experiment 3, we measured six negative self-conscious emotions (*guilt, shame, embarrassment, disgrace, remorse, sorry*) and two anxiety-related emotions (*worry, anxiety*). The ninth emotion (*envy*) was used as a control item.³ Factor analysis suggested one meaningful factor that explained 95% of the variance of the items. Eight items loaded on this factor, which we labeled *negative emotions* (α = .90, M = 6.56, SD = 1.82).⁴ Envy (M = 1.26, SD = 1.52) loaded below .40 on this factor and was thus dropped.

*Ingroup reputation concern.* A subtle measure of ingroup reputation concern followed, consisting of five items. An open-ended question asked participants why they would feel the reported emotions. After writing a response, participants completed three sentences: “People who saw the incident may form negative opinions about ...”, “My behavior may reflect poorly on ...”, “I would feel responsible for possibly causing troubles to...”. Finally, a second open-ended question asked “What kind of negative effects might your behavior in the story have had?” All questions could be answered referring to the self or the ingroup. For each of these five items, two independent coders judged whether or not participants referred to the U.S. (ingroup) in their answers. If participants referred to the U.S., it was coded 1, if not it was coded zero. The sum of the resulting five codings constituted an index of participants’ concern over the ingroup (M =
0.91, \( SD = 1.40 \)). Four participants also mentioned that their behavior could reflect poorly on women (two in the intragroup and two in the intergroup condition).

Participants then answered the same, more blatant ingroup reputation concern questions as in Experiment 3 (\( \alpha = .90, M = 3.56, SD = 2.02 \)). This explicit measure of ingroup reputation concern was strongly correlated with the subtle measure, \( r = .52, p < .001 \). Factor analysis showed that the items of the explicit and the subtle measure loaded onto the same factor. The explicit and subtle group reputation concern scores were thus each standardized and then averaged into a composite score of ingroup reputation concern (\( M = 0.00 \) and \( SD = 0.87 \)).

**Appraised cause of emotions.** We again assessed emotion attribution with two items, but here we focused on parallel assessment of the transgression’s consequences for the ingroup’s reputation (“I would feel bad because my behavior may reflect poorly on other Americans”; \( M = 3.72, SD = 2.42 \)) and for the individual reputation (“I would feel bad because my behavior may reflect poorly on me”; \( M = 5.96, SD = 1.28 \)). Participants’ answers were recorded on continuous scales from 0 (Not at all) to 7 (Very much).

**Perceived ingroup entitativity.** Perceived entitativity of the U.S. was assessed with five items adapted from past research (“Some groups have the characteristics of a ‘group’ more than others do. To what extent does the United States qualify as a group?”, “How cohesive of a group is the United States?”, “Overall, how similar do you think are Americans to each other?”, “To what extent do you think Americans possess some underlying common characteristics?”, “To what extent do you think Americans have a sense of common fate?”). The answers were rated on continuous scales from 0 (Not at all) to 7 (Very much), \( \alpha = .81, M = 3.96, SD = 1.19 \).
Manipulation check. The manipulation check asked participants to recall whether the location of the argument described in the scenario happened in a foreign country or in Southern California.

Results

We ran moderated regression analyses with condition as categorical IV, entitativity (centered, not affected by condition, $p = .206$) as a continuous moderator, and with negative emotions, ingroup reputation concern and attributed cause of emotions as DVs.

Emotions. The main effect of condition on negative emotions was marginally significant, $F(1, 186) = 2.83, p = .094, \eta^2 = .015$, consistent with the finding of Experiment 3. Entitativity also predicted negative emotions, $F(1, 186) = 9.58, p = .002, \eta^2 = .049, \beta = .41$, with higher perceived entitativity predicting more negative emotions. Most importantly, the predicted interaction between entitativity and condition was significant, $F(1, 186) = 5.85, p = .017, \eta^2 = .031$. Entitativity significantly influenced negative emotions in the intergroup condition, $t(186) = 4.67, p < .001$, but not in the ingroup condition, $t(186) = .042, p = .676$. From another angle, participants who perceived the U.S. as high in entitativity (+1 SD) felt significantly more negative emotions when collective responsibility was present ($M = 7.52$) rather than when it was absent ($M = 6.45$), $t(186) = 2.96, p = .004, d = .43$. Participants who viewed the U.S. as low in entitativity (-1SD), on the other hand, felt similar levels of negative emotions when collective responsibility was present ($M = 6.06$) rather than when it was absent ($M = 6.28$), $t(186) = -.58, p = .562$ (see Figure 3). Envy, the control emotion, was not affected by condition, entitativity, or their interaction, $Fs < 0.70, ps > .400$.

Ingroup reputation concern. Analyses showed significant main effects of
condition, $F(1, 186) = 174.17, p < .001$, $\eta^2 = .48$, and entitativity, $F(1, 186) = 11.22, p = .001, \eta^2 = .057$. Group reputation concern was significantly higher in the intergroup ($M = 0.56$) than in the intragroup condition ($M = -0.61$), and the greater the entitativity, the greater the perceived damage to the ingroup’s image ($\beta = .16$). The interaction effect did not reach significance but showed a trend, $F(1, 186) = 2.70, p = .102, \eta^2 = .014$, and the simple effects were in line with predictions. Entitativity significantly influenced concern in the intergroup condition, $t(186) = 4.23, p < .001$, but not in the intragroup condition, $t(186) = 1.06, p = .292$. From another angle, participants who perceived the U.S. as high in entitativity (+1 SD) were significantly more concerned when collective responsibility was present ($M = .81$) rather than when it was absent ($M = -.54$), $t(186) = 10.47, p < .001$. Participants who viewed the U.S. as low in entitativity (-1SD) were also more concerned when collective responsibility was present ($M = .34$) rather than when it was absent ($M = -.70$), but to a lesser extent, $t(186) = 7.78, p < .001$.

Attributed cause of emotions. Personal cause (i.e., attributing the experienced emotions to the personal consequences of the behavior) was not affected by condition, entitativity, or their interaction, $ps > .149$. People attributed emotions to damage of their personal image to the same extent in the intergroup ($M = 5.98$) and the intragroup condition ($M = 5.93$). The group cause (i.e., attributing the experienced emotions to the group level consequences of the behavior), on the other hand, was affected by condition, $F(1, 186) = 45.45, p < .001, \eta^2 = .195$, and by entitativity, $F(1, 186) = 17.19, p < .001, \eta^2 = .085$. It was higher in the intergroup ($M = 4.63$) than in the intragroup condition ($M = 2.70$), and was positively predicted by entitativity ($\beta = .66$). Importantly, these effects were qualified by a marginally significant two-way interaction between condition and
entitativity, $F(1, 186) = 2.77, p = .098, \eta^2 = .015$. In the intragroup condition, entitativity did not predict group cause, $t = 1.54, p = .125$, whereas in the intergroup condition it did, $t(186) = 4.92, p < .001, d = .72$. Participants who perceived the U.S. as high in entitativity (+1 SD) attributed their emotions significantly more to group level consequences when collective responsibility was present ($M = 5.64$) rather than when it was absent ($M = 3.06$), $t(186) = 5.94, p < .001$. Participants who viewed the U.S. as low in entitativity (-1SD) also attributed their negative emotions to the group level consequences when collective responsibility was present ($M = 3.78$) rather than when it was absent ($M = 2.26$), although to a lesser extent, $t(186) = 3.37, p < .001$.

*Mediation analyses.* Given that perceived entitativity moderated condition effects on negative emotions and on ingroup reputation concern, we conducted a moderated mediation analysis, with condition as IV, entitativity as a moderator, ingroup reputation concern as mediator, and negative emotions as the DV, with 95% bias-corrected confidence intervals and 5000 bootstrap resamples (Hayes, 2012, model 59). The conditional confidence intervals of the indirect effect included zero ($CI_{95} = [-.048, .526]$) at low levels of entitativity, but not at high levels of entitativity ($CI_{95} = [.022, .863]$), indicating that ingroup reputation concern mediated the condition effect among those who perceived the U.S. as highly entitative.\(^6\)

**Discussion**

The results conceptually replicated the findings of Experiment 3 with a different manipulation and with a new outcome variable. Most importantly, they extended Experiment 3 and confirmed our hypothesis that entitativity moderates collective responsibility processes within the individual actor as well. Even when collective blame
was possible, only wrongdoers who perceived the ingroup as highly entitative felt more negative emotions for a past wrongdoing. Importantly, participants attributed their feelings to hurting their personal image at the same level in the two conditions, but attributed it significantly more to hurting the ingroup’s image in the intergroup condition. Also, similar to Experiment 3, we showed that the condition effect on negative emotions was mediated by ingroup reputation concern.

In the first two studies, we showed that people regulated their behavior when it could impact the reputation of their group. Of course, people are not always perfectly aware of potential intergroup contexts or able to completely control their behavior. In these instances, even if their behavior is not perfect, when people realize *afterwards* that their behavior could trigger collective blame, collective responsibility processes impact how badly people feel for their behavior. We hypothesized that this after-the-fact response is not merely emotional, but should also influence subsequent behavior to better protect the ingroup’s reputation from harm. Experiment 5 thus tested whether the negative emotions felt over activating the ingroup’s collective blaming motivate people to behave better in the *future*.

**Experiment 5**

We showed in Experiments 3 and 4 that people feel more negatively about a past wrongdoing if it could activate collective blame of the ingroup. The question is whether collective responsibility processes would induce more positive behavior even in this case, when the wrongdoing has already happened, or whether people would become indifferent or resigned to the past harm they caused. To answer this question, we used a modified dictator game paradigm.
In this study, participants played two rounds of a dictator game as the allocator. After the first round of the game participants received bogus feedback about how selfishly they had behaved. Additionally, we manipulated whether this allegedly selfish behavior could or could not damage the ingroup’s reputation – while keeping the individual actor anonymous to control for possible repercussions for their individual reputation. Then participants played a second round of the game. We predicted that people would experience more negative emotions when their alleged selfishness could hurt the ingroup’s reputation in the eyes of an outgroup and that this increased level of negative emotions would lead to less selfish behavior in the second round of the game (see Figure 4).

Experiment 5 served also as an important replication of Experiments 3 and 4. Those studies used vignettes in which people imagined a scenario and reported their expectations of their emotional responses to the scenario. While this approach allows considerable control over the event being judged and there is empirical evidence that anticipated emotions and appraisals in hypothetical scenarios are often very consistent with the emotions and appraisals reported in real situations (Robinson & Clore, 2001), it was important to corroborate the validity of our results and extend our findings with behavioral situations. The dictator game allowed us to do so. Furthermore, if the emotions were to have behavioral consequences, it would render additional evidence that the reported emotions were indeed experienced (rather than just reported out of demand characteristics).

Method

Participants. Seventy-seven college undergraduates ($M_{age} = 19.73, SD = 1.26,$
range 18-23 years; 66 female) participated for research credit as well as for a $2 monetary reward.

**Materials and procedure.**

Upon arrival, the experimenter led participants into a testing room and seated them at a computer that presented the survey and instructions. After giving consent, participants completed a measure of perceived entitativity of their university’s student body.

**Entitativity.** We adapted five items from a perceived group entitativity scale (Ip, Chiu, & Wan’s, 2006). Participants answered questions regarding the ingroup’s perceived cohesiveness (e.g. “To what extent you think that others perceive the university’s students as a cohesive real group?”) and homogeneity (e.g. “Overall, how similar do you think are the university’s students to each other?”). The order of the items was randomized. The items were measured on continuous scales from 1 (Strongly disagree) to 9 (Strongly agree). They loaded on one single factor ($M = 5.45, SD = 0.93, \alpha = .69$).

**Dictator game.** Participants were assigned to the role of allocator and told to play the game with another player who was ostensibly sitting in another room. In reality there was no other player. Participants were under the impression of playing the game only once and did not know in advance that they would be asked to play the game a second time. First, participants were given detailed instructions on how to play the game. The experimenter tested for participants’ understanding of the game and walked them through the rules again. Participants were then given twenty aluminum coins, each worth five cents, to anonymously divide between themselves and the other player. Participants were
informed that there was no danger of retaliation from the other player, that they would lose whatever they chose to give to the other player, and that at the end of the study they would get real money for the coins they had kept for themselves. The coins that participants kept for themselves had to be inserted into a digital coin acceptor machine connected to the computer. The coins that participants wanted to give to the other player had to be placed into an envelope. Participants were told that the computer automatically counted the coins inserted into the coin acceptor machine and therefore their decision would remain unknown to the experimenter. The experimenter then left the room and participants distributed the coins. Then the experimenter returned, took the envelope and ostensibly brought it to the other player. In reality, outside the testing room the experimenter labeled and filed the envelope for later data entry. Participants meanwhile continued the survey.

*Induction of negative emotions.* Following the coin allocation, the survey instructions informed participants in both conditions that the game that they had just played measures moral character based on how many coins they had given to the other player. They were also told that based on the number of coins they had kept for themselves, the computer had calculated that they gave four coins less to the other player than others gave on average. As every participant received this message, we kept constant the moral transgression across conditions – together with the possible emotional consequences due to the nature of their actions per se.

*Manipulation of collective responsibility.* To manipulate collective responsibility, in the experimental condition the instructions additionally mentioned that the study was part of a regional inter-college comparison assessing students’ moral character in the
region. They were told that the computer had randomly selected their data to compare the two in-town colleges in moral character. To play down any personal accountability, participants were assured that this comparison was completely anonymous and only their college’s name would appear next to their data, but not their personal names. Again, in both conditions participants were led to believe that they had behaved selfishly. In the experimental condition, however, this individual selfish behavior could also reflect on the individual’s ingroup. We did not specify what the average was that others gave to the second player, only that participants had given four coins less than this average. Regardless of the actual number of coins that participants gave to the other player, almost everybody could believe that they behaved selfishly. Based on a recent meta-analysis, only 6.5% of people give away more than 80% of their endowment in dictator games (Engel, 2010). For the few participants who were highly generous and gave away more than 16 of their twenty coins, the feedback was illogical and thus meaningless and confusing. In our study, out of 77 participants, seven (9.1%) gave more than 16 coins and were thus excluded from the analyses.

*Emotions.* Participants reported how they felt about the (first round of the) game. Eight items measured negative emotions (*guilt, remorse, disgrace, shame, sorry, worry, anxiety, anger*), and four items measured positive emotions (*pleasure, satisfaction, contentment, joy*) for control purposes and because participants may have felt good about earning money. All emotions were rated on continuous scales from 1 (Not at all) to 9 (Very much). Factor analyses suggested two meaningful factors. One of the factors comprised the eight negative items (*round 1 negative emotions* [i.e. negative emotions felt after round 1]; $M = 4.05, SD = 1.70, \alpha = .94$), the other the four positive emotions
(round 1 satisfaction [i.e. satisfaction felt after game round 1]; \( M = 4.36, SD = 1.41, \alpha = .79 \)).

**Ingroup reputation concern.** Participants’ appraisal of the consequences of their actions for the ingroup’s reputation was assessed with four statements (e.g. “I am afraid that my decision in the game damaged <participants’ university’s> reputation”, “I worry that the way I played the game could paint a poor image about <participants’ university> students’ moral character”). Participants expressed their agreement with the statements on continuous scales from 1 (Not at all) to 9 (Very much). The four items formed one factor labeled *round 1 ingroup reputation concern* (\( M = 3.31, SD = 2.03, \alpha = .97 \)).

**Game round 2.** Participants were then informed that they would play the game again, and again as the allocator. They were walked through the rules again and got a new set of twenty coins. After distributing all the coins by putting them into the coin acceptor machine or into the envelope, they notified the experimenter, who took the sealed envelope and left the room. In the hallway the experimenter labeled and filed the second envelope. Participants meanwhile continued the survey. They were told that the computer would give them feedback about their second game’s performance later on, but in reality this feedback never came. Participants completed the same emotion and ingroup reputation concern measures they had completed after the first round; factor analyses suggested the same factors as for the round 1 measures. Composite scores were computed by averaging the eight negative emotion items into *round 2 negative emotions* (\( M = 2.90, SD = 1.41, \alpha = .92 \)), the four positive emotions into *round 2 satisfaction* (\( M = 5.32, SD = 1.11, \alpha = .73 \)), and the ingroup reputation concern items into *round 2 ingroup reputation concern* (\( M = 3.57, SD = 2.25, \alpha = .97 \)).
At the end, participants completed basic demographic questions and were asked to describe what the study’s goal was. Regardless of their behavior in the game, participants received the maximum possible payout of $2. After participants were debriefed and left, the experimenter opened the envelopes and recorded the number of coins participants had given to the other player in the first envelope (round 1 coins) and in the second envelope (round 2 coins) as (behavioral) dependent variables.

Results

As explained above, seven participants were excluded because they gave more than 16 coins in the first game and thus the bogus feedback was not credible. Four participants did not follow the instructions (one did not notify the experimenter during the game when told to do so, and three decided to terminate the study before it was finished), and three participants suspected that there was no other player. It is not uncommon that utilizing bogus feedback and bogus partners leads to suspicion about the authenticity of the feedback and the existence of the partners (e.g. Galinsky, Magee, Gruenfeld, Whitson, & Liljenquist, 2008; Mojzisch, Kerschreiter, Faulmüller, Vogelgesang, & Schulz-Hardt, 2014; Muller & Butera, 2007). That left 63 participants in the sample for subsequent analysis.

In a failure of random assignment, entitativity was higher in the experimental ($M = 5.83$) than in the control condition ($M = 5.08$), $F(1, 61) = 11.80, p = .001$ (although it had been measured before the manipulation). To make sure that this failure of random assignment did not account for the results, we performed every analysis reported below with entitativity as a covariate as well. All results reported remained significant when we did so.
The effects of collective responsibility on ingroup reputation concern and negative emotions. To answer our primary question of how collective responsibility affects ingroup reputation concern and negative emotions, we introduced ingroup reputation concern and negative emotions as DVs and condition as IV into general linear models. As predicted, round 1 ingroup reputation concern was significantly higher in the experimental ($M = 4.46$) than in the control condition ($M = 2.19$), $F(1, 61) = 28.26, p < .001, \eta^2 = .32$. Round 2 ingroup reputation concern was also higher in the experimental ($M = 4.93$) than in the control condition ($M = 2.26$), $F(1, 61) = 34.19, p < .001, \eta^2 = .36$. Similarly, participants felt significantly more negative emotions in the experimental ($M = 4.56$) than in the control condition ($M = 3.56$) after the first game (round 1 negative emotions), $F(1, 61) = 5.77, p = .019, \eta^2 = .09$. Participants also felt more negative emotions in the experimental ($M = 3.44$) than in the control condition ($M = 2.42$) after the second game (round 2 negative emotions), $F(1, 61) = 8.36, p = .005, \eta^2 = .12$. The manipulation did not affect the satisfaction participants reported after the first or the second game, $Fs < .33, ps > .56$.

The mediating role of ingroup reputation concern. Next we tested whether perceived ingroup reputation concern mediated the condition effect on the negative emotions, using 95% bias-corrected confidence intervals and 5000 bootstrap resamples (Hayes, 2012, model 4). We introduced condition as IV, round 1 ingroup reputation concern as mediator, and round 1 negative emotions as DV. Condition significantly predicted round 1 ingroup reputation concern, $t = 5.31, p < .001$, which in turn predicted round 1 negative emotions, $t = 4.84, p < .001$. The confidence interval of the indirect effect did not include zero ($CI_{95} = [.368, 1.042]$), indicating a significant indirect effect.
Unlike in Experiment 2, where the number of coins predicted negative emotions, in Experiment 5 we did not expect that round 1 coins would predict emotions because we gave every participant the same (false) feedback about how selfishly they had behaved. As predicted, the number of coins that participants gave to the other player did not predict negative emotions or ingroup reputation concern after the first game.

The effect of ingroup reputation concern and negative emotions on future behavior. Next we tested how collective responsibility, ingroup reputation concern and negative emotions after game 1 influenced participants’ behavior in game 2. First we ran a repeated-measures ANOVA with the number of coins people gave away in the first game (round 1 coins) and in the second game (round 2 coins) as DVs, condition as IV, and game round as the repeated-measures factor. The condition by game round interaction was significant, \( F(2, 61) = 5.91, p = .018 \), indicating that the effect of condition on behavior in round 1 was different from the effect of condition on behavior in round 2. The subsequent univariate analysis showed that, as predicted, only round 2 coins was significantly affected by condition, \( F(1, 61) = 6.43, p = .014, \eta^2 = .10 \), but not round 1 coins, \( F(1, 61) = 2.71, p = .105, \eta^2 = .04 \).

The mediating role of negative emotions. To fully uncover how collective responsibility shapes subsequent behavior and to see what role emotions play in it, we ran a mediational analysis with 95% bias-corrected confidence intervals and 5000 bootstrap resamples (Hayes, 2012, model 4) with condition as IV, round 1 negative emotions as mediator, round 2 coins as DV, and round 1 coins as covariate. Condition significantly predicted round 1 negative emotions, \( t = 2.53, p = .014 \), which, in turn, marginally significantly predicted round 2 coins, \( t = 1.85, p = .069 \). The direct effect became non-
significant, \( t = 1.24, p = .22 \), and the confidence interval of the indirect effect did not include zero (CI\(_{95} = [0.064, 1.00]\)), indicating a significant indirect effect.

*The effect of future behavior on subsequent negative emotions.* Finally, we tested whether the more positive behavior in the second game would alleviate negative emotions after the second game. We ran a sequential mediation (Hayes, 2012, model 6) with condition as IV, round 1 negative emotions and round 2 coins as mediators, round 2 negative emotions as DV, and round 1 coins as covariate. This analysis reflects the hypothesized process, in which condition (collective responsibility) influences the negative emotions felt after the first game and where the negative emotions in turn influence how many coins people gave away in the second game—a behavior that eventually reduced the negative emotions experienced after the second game. Supporting our hypothesis, condition significantly influenced round 1 negative emotions, \( t = 2.52, p = .014 \), which, in turn, marginally significantly influenced round 2 coins, \( t = 1.85, p = .069 \). Round 2 coins, in turn, marginally significantly influenced round 2 negative emotions, \( t = -1.79, p = .079 \). The direct effect became non-significant, \( t = 1.40, p = .166 \), and the confidence interval of the indirect effect for the full model (condition > round 1 negative emotions > round 2 coins > round 2 negative emotions) did not include zero (CI\(_{95} = [-.127, -.001]\)), indicating a significant indirect effect (see Figure 5). Of note, as we predicted, in the collective blame condition negative emotions about one’s selfish behavior predicted more generous behavior in the next game, and the more generous behavior then led to less negative emotions. In the alternative models that did not include round 1 negative emotions or that reversed the order of round 1 negative emotions and round 2 negative emotions we did not find significant indirect effects.
Discussion

Experiment 5 replicated the findings of Experiment 3 and 4 in the laboratory by using an economic game paradigm. Participants felt more negative emotions when they learned (after the fact) that their selfish behavior could have caused collective blame for the ingroup. This increase in negative emotion led participants to behave more generously in the next dictator game. Being generous in the second game then successfully reduced the level of experienced negative emotions. Importantly, participants remained anonymous and their individual reputation was not threatened by their behavior. The bogus feedback about their selfishness was also constant across conditions to control for moral inferences about the self.

General Discussion

By synthesizing the ingroup and outgroup perspectives of collective responsibility, we concluded that there is a third perspective that had yet to be understood. This third perspective is that of the individual actor whose behavior triggers the attribution of collective blame. The results rendered converging evidence that collective responsibility processes influence the individual actor as well. When collective blame was possible, people behaved more prosocially (Experiments 1 and 2)—even against their own personal interest. Similarly, for past behavior, when collective blame was possible people appraised their behavior in terms of its (potential) effects on the ingroup’s reputation. People experienced more ingroup reputation concern and negative emotions for the same past behavior when they thought that it could reflect negatively on the ingroup in the eyes of outgroups (Experiment 3-5). The increased level of negative emotions then led to subsequent behavior aimed to improve the ingroup’s image, even at
the actor’s personal expense (Experiment 5). These effects of collective responsibility on appraisals and emotions were further moderated by perceived ingroup entitativity (Experiment 4). The fact that entitativity matters in all three perspectives of collective responsibility (i.e., outgroup, ingroup, actor) as a moderator corroborates that the seemingly distinct psychological mechanisms within the three perspectives (e.g. collective guilt, person to group generalization) indeed inform different perspectives of the same psychological phenomenon—that is, collective responsibility. The converging evidence for our proposed third perspective of collective responsibility and its underlying psychological processes is quite robust given that the five experiments used different paradigms (imagined responses to a vignette vs. behavior in experimental games), experimental settings and data collection strategies (online vs. laboratory), as well as types of groups (national vs. college identity vs. minimal group).

Integrating all three perspectives of collective responsibility, it becomes clear that collective responsibility has interconnected but different meanings in the three perspectives. For the outgroup, collective responsibility means blaming the members of the actor’s entire group collectively for his or her actions and in some circumstances extending explicit or implicit punishment to people who were not, in a literal sense, the cause of the blameworthy action. For the ingroup, collective responsibility means feeling responsible for the actions of individual group members and experiencing self-conscious emotions for actions committed by others who are tied to the self through group membership. For the individual actor, collective responsibility means feeling responsible for the collective that the individual belongs to and being attuned to both the potential for outgroups to blame the ingroup. The individual actor’s perspective that we have
developed in the current research has the potential to inform these prior lines of research. It is possible, for example, that when the individual actor shows remorse for his wrongdoing, the outgroup would attribute less collective blame to his entire group. Likewise, the ingroup’s reaction to collective blame might depend on the individual actor’s response as well. If it is true that the ingroup’s expression of collective shame functions to appease the outgroup (Giner-Sorolla, Castano, Espinosa, & Brown, 2008), the individual wrongdoer’s expression of shame should lessen the need for the ingroup to do the same.

The individual actor’s perspective also connects with research on stereotype threat. The current research shows that the understanding that one’s individual behavior can influence the ingroup’s reputation in the eyes of the outgroup goes beyond confirming negative stereotypes. Group reputation concern (Shapiro & Neuberg, 2007) is thus a more general phenomenon. Nonetheless, we would expect that the actor’s concern over collective blame would likely be strengthened if their actions could confirm a pre-existing negative reputation or stereotype of the group. However, given our findings from the minimal group paradigm, pre-existing stereotypes are not required.

**The self-regulatory function of collective responsibility and the intergroup context**

The individual actor’s perspective of collective responsibility also reveals how collective responsibility processes can regulate individual behavior. One of the responsibilities of a good group member is to not bring blame and disrepute to the group. Even if a person were not to refrain from a negative action (e.g. selfish behavior) out of concern for the self or the harm the action might cause, she may still do so for the sake of the ingroup, as long as there is a possibility that other groups would blame the ingroup
because of the person’s individual action. In this sense, collective responsibility can be viewed as a source of *group-based self-regulation*, where anticipated or actual group level consequences guide individual behavior. Self-regulation is generally seen as the basis of proper conduct and peaceful social life (Baumeister, Schmeichel, & Vohs, 2007; Finkel & Campbell, 2001; Tangney, Baumeister, & Boone, 2004; Vohs & Ciarocco, 2004). Recognizing the possibility to hurt the ingroup’s reputation and the motivation to avoid it may provide additional regulatory resources to think and act properly.

Research on self-regulation has shown that understanding what is the right course of action and being motivated to execute it does not guarantee that people will actually do so. Similarly, understanding the possible negative group level consequences of one’s individual behavior provides additional, but not necessarily sufficient, self-regulatory power to do the right thing. While the recognition that one’s behavior may hurt the ingroup does not guarantee that people will control their behavior, even if a person has failed to behave properly and harmed the ingroup’s reputation, collective responsibility and its power of group-based self-regulation may induce proper conduct in the future.

Remarkably, collective responsibility exercises its self-regulatory power specifically in *intergroup* situations and relations—an area of social life that is widely seen as being prone to negative biases and destructive behavior. Intergroup contexts have frequently been conceptualized as a source of animosity, conflict, and violence that, compared to intragroup contexts, has detrimental effects on behavior (e.g., Wildschut, Pinter, Vevea, Insko, & Schopler, 2003). However, in line with an emerging interpretation of the intergroup context as a possible source of positive effects (e.g. Pittinsky, 2012; Spears, 2010), we have shown that the intergroup context can promote
positive behavior through collective responsibility processes. For the individual actor, intergroup contexts carry the extra responsibility of being a good representative of the ingroup. Even without formally acting on behalf of the ingroup, people perceive themselves and others in intergroup contexts as if they were representing their respective groups. This representative role requires people to manage the impression of their ingroups, extending the chore and scope of impression management beyond a purely personal concern. People are thus motivated to behave more prosocially in the intergroup context.

**The feeling of hurting the ingroup’s image**

We have found that people felt negative emotions for tarnishing the ingroup’s image and that these emotions in turn motivated behavior that could improve the ingroup’s image. We assessed the individual actor’s emotional experiences by measuring guilt and shame related emotions, and in each study they formed one single factor. Although guilt and shame can be empirically distinguished, much prior research shows them to be correlated (e.g., Schmader & Lickel, 2006b) and have many underlying appraisals in common (Smith & Ellsworth, 1985). Nonetheless, in future research there may be value to conduct studies designed to teasing apart the particular impact of feeling guilt versus shame for bringing collective blame to one's group. Past research has found that shame is commonly more strongly related to the perceived failure of the self and damaged self-image, while guilt is commonly triggered by personal responsibility for a specific negative outcome (Baumeister, Stillwell, & Heatherton, 1994; Deonna & Teroni, 2008; Eisenberg, 2000; Gausel & Leach, 2011; Tangney & Fischer, 1995; Wohl & Branscombe, 2010). The extent to which people feel shame versus guilt has also been
found to predict behavior after wrongdoing. Thus, future research might test these fine-grained distinctions on the actor's response to bringing collective blame to the ingroup.

While guilt is more strongly associated with reparation and shame with withdrawal (Tangney, Miller, Flicker, & Barlow, 1996), both guilt and shame can motivate compensation in intergroup contexts (Brown & Cehajic, 2008; Brown, González, Zagefka, Manzi, & Cehajic, 2008; Gausel & Leach, 2011; Gausel, Leach, Vignoles, & Brown, 2012). It is thus probable that the individual actor’s self-conscious emotions lead to positive behavior via compensatory motives. In this case, feeling guilty and ashamed over damaging the ingroup’s image motivates the actor to repair the image. One of the most important findings about the ingroup’s reaction to collective blame is how collective guilt and shame felt in collective responsibility processes motivate compensatory behavior indirectly, through the wrongdoer’s group members (Doosje et al., 1998; Iyer et al., 2007). Complementing this finding, our results suggest that collective responsibility might motivate compensation directly as well, through the wrongdoer him- or herself (Experiment 5).

**Boundary conditions and future directions**

In the present set of studies participants were keenly aware that the individual behavior could hurt the ingroup’s reputation. In many real life situations, however, the situation as well as the relationship between individual behavior and its group level consequences can be ambiguous or indirect, and thus the right course of action uncertain. One reason of this ambiguity is that people often erroneously judge what is good or bad for the ingroup and have a tendency to overestimate how much their personal opinion is representative of their group in general (Ross, Greene, & House, 1976). Being convinced
that the ingroup should display toughness, for example, individual actors may not shy
away from violence out of the false belief of serving the ingroup’s ideal reputation.

Ingroup heterogeneity and the existence of subgroups may also contribute to
ambiguity. Based on their political beliefs, for example, American liberals and
conservatives may perceive very different behaviors as suitable to promote a positive
image of the U.S. in the world (e.g. with respect to U.S. foreign policy). A related
possibility is that the ingroup or some of its members are in fact hostile toward the
outgroup and would back the actor. In such cases the collective responsibility processes
we have demonstrated would not occur, because the actor would not be concerned about
the outgroup’s possible reaction given the ingroup's support. In fact, rather than behaving
more prosocially, in situations characterized by intergroup hostility people may
purposefully behave negatively in order to provoke negative group level consequences.
Thus, it is important to find out how collective responsibility processes influence the
individual actor under more ambiguous circumstances.

Specific intergroup factors may chronically activate collective responsibility
processes within the individual actor. First, the nature of pre-existing intergroup relations
between the ingroup and outgroup may moderate collective responsibility processes in
the individual actor. In intergroup conflict, where the context emphasizes intergroup
division, individuals cannot avoid being perceived as a representative of their group in
the eyes of the adversary outgroup (Haslam, Oakes, McGarty, Turner, & Onorato, 1995;
Smith & Leach, 2004; Tajfel & Turner, 1979; Turner et al., 1987), and we thus expect the
identified effects to be amplified. If the ingroup as a whole is attempting to repair
relations with the outgroup, then bad actions by individual group members should
provoke particularly strong reactions. But, as we discussed earlier, if there is antagonism then actor's aggressive behavior may well be supported or even celebrated. In harmonious intergroup contexts, on the other hand, intergroup boundaries may be less pronounced, and therefore people’s perception that they represent the ingroup in front of an outgroup as well as any ensuing motivational and emotional consequence should be reduced regardless of whether the actor does or does not expect ingroup concern for his action.

Second, the group’s relative power could also influence collective responsibility processes in the individual actor. People in lower power groups are more concerned with social disapproval and norm violation (Emerson, 1962; Galinsky et al., 2008). Therefore, being a member of a relatively less powerful group (or a low power person within a group) should increase the sensitivity to, and therefore the effects of, collective responsibility processes. These factors suggest that the individual actor’s perspective should be of special interest among minority members who are often taught, explicitly and implicitly, that they represent their group with every single step they take.

**Concluding remarks**

Psychology has long been investigating collective responsibility and its processes, showing that individual behavior has group level consequences in that it affects ingroup and outgroup members. Introducing a third perspective, that of the individual actor who activates collective responsibility processes, we have shown that individual actors are aware of the group level consequences, too. Thus they try to not bring blame and disrepute to the group; when they fail, they are concerned, feel bad about it, and will try to remedy it in the future. By learning from this new perspective how collective
responsibility processes guide individual behavior, we may be able to leverage collective responsibility to improve people’s behavior even in the intergroup context.
Notes

1 In all of the studies we measured emotions first, before measuring ingroup reputation concern. Although in the psychological process the appraisal of the events precedes the emotions felt about them, it was important to measure participants’ emotions while participants were most naïve about the studies’ goal of investigating collective responsibility processes. Asking first about ingroup reputation concern could have distorted responses and possibly led to demand characteristics on the emotion measure.

2 Introducing the three types of emotions (negative emotions, concern, unrelated) as a repeated factor in the ANOVA did not yield a significant condition by emotion type interaction. Scores on the unrelated words showed a skewed pattern even after eliminating outlier values that were more than 2.5 standard deviations from the mean. The skewness was to be expected, as most participants reported almost no nostalgia or envy. We thus transformed the scores of the unrelated emotions. Inverse transformation (Tabachnik & Fidell, 2007) reduced skewness from 2.03 to -0.68. Entering the transformed scores into the repeated measures ANOVA resulted in a significant interaction effect between condition and emotion type, $F(2, 112) = 3.66, p = .029$, where only negative emotions was significantly affected by condition.

3 We used the same list of emotions as in Experiment 3 except dropping one of the control emotions (nostalgia).

4 Using the same 6-item negative emotions factor as in Experiment 3 produced virtually the same results in every analysis.

5 Introducing type of emotions as a repeated factor resulted in a marginally
significant interaction effect between condition, entitativity and emotion type, \( F(1, 185) = 3.04, p = .083 \). Similar to the unrelated emotions in Experiment 3, the scores on the envy item were skewed (1.49) and showed little variation around the minimal value. Transformation reduced skewness (-0.01). Use of the inverse transformed score (Tabachnik & Fidell, 2007) the repeated measures ANOVA resulted in a significant interaction between type of emotions, condition and entitativity, \( F(1, 185) = 5.08, p = .025 \).

6 We also replicated the simple mediation as in Experiment 3.

7 Compared to Experiment 4, we replaced embarrassment with anger.

8 Round 1 coins was not significantly affected by condition but showed a trend in the same direction as round 2 coins. While the reported repeated-measures ANOVA is the most stringent test to reveal differential condition effects on round 1 coins and round 2 coins, because of the univariate trend on round 1 coins we performed additional analyses on round 2 coins, using round 1 coins as a covariate or full factor. The condition effect on round 2 coins remained marginally significant regardless of whether we entered round 1 coins as a covariate, \( F(1, 60) = 3.51, p = .066 \), or as a full factor, \( F(1, 59) = 3.66, p = .061 \), in the ANOVA.
References


Who else was to blame for the Columbine high school shootings? Personality and Social Psychology Bulletin, 29, 194–204.


Emerging directions in research. Social And Personality Psychology Compass, 5, 153-163.

Subjective group dynamics and derogation of in-group deviants when generic norms are undermined. Journal of Personality and Social Psychology, 81, 436-447.


Figure 1: Negative emotions felt over individual behavior as the function of collective blame possibility and the prosocial nature of the behavior.
Figure 2: Ingroup reputation concern mediates the effect of collective blame activation on negative emotions. * $p < .05$. 

95% confidence interval [.021, .318]
Figure 3: Negative emotions felt over an individual transgression as the function of social context and perceived ingroup entitativity.
Figure 4: Experiment 5: The timeline of the experimental procedure.
Figure 5: Sequential mediation: Collective blame activation increases negative emotions; negative emotions, in turn, increase compensatory behavior; compensatory behavior reduces subsequent negative emotions. *p < .05.