



On the near miss in public good dilemmas: How upward counterfactuals influence group stability when the group fails

David de Cremer, Eric van Dijk

► To cite this version:

David de Cremer, Eric van Dijk. On the near miss in public good dilemmas: How upward counterfactuals influence group stability when the group fails. *Journal of Experimental Social Psychology*, 2010, 47 (1), pp.139. 10.1016/j.jesp.2010.09.006 . hal-00918798

HAL Id: hal-00918798

<https://hal.science/hal-00918798>

Submitted on 15 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

On the near miss in public good dilemmas: How upward counterfactuals influence group stability when the group fails

David De Cremer, Eric van Dijk

PII: S0022-1031(10)00212-X
DOI: doi: [10.1016/j.jesp.2010.09.006](https://doi.org/10.1016/j.jesp.2010.09.006)
Reference: YJESP 2538

To appear in: *Journal of Experimental Social Psychology*

Received date: 10 November 2009
Revised date: 26 August 2010



Please cite this article as: De Cremer, D. & van Dijk, E., On the near miss in public good dilemmas: How upward counterfactuals influence group stability when the group fails, *Journal of Experimental Social Psychology* (2010), doi: [10.1016/j.jesp.2010.09.006](https://doi.org/10.1016/j.jesp.2010.09.006)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Running Head: ON THE NEAR MISS IN PUBLIC GOOD DILEMMAS

On the Near Miss in Public Good Dilemmas:
How Upward Counterfactuals Influence Group Stability When the Group Fails

David De Cremer
Rotterdam School of Management, Erasmus University

Eric van Dijk
Leiden University

Address correspondence to David De Cremer, Centre of Behavioural Ethics
(www.erim.nl/behaviouralethics), Rotterdam School of Management, Erasmus
University, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands, Fax: 00 31 10
4089012, Email: ddecremer@rsm.nl.

Abstract

In two studies we investigated the impact of degree of collective failure in a public good dilemma (near miss vs. large miss) on group members' negative reactions (negative affect, attributions of responsibility for the failure, and intention to leave the group). The results show that upward counterfactual thinking has more impact on members' negative responses when experiencing a near miss rather than a large miss. In Experiment 1, the results show that in the case of a near miss (and not a large miss), negative affect and attributions of responsibility were higher when other-focused counterfactuals rather than self-focused counterfactuals were elicited. Negative affect was found to mediate the effect on attributions of responsibility. Experiment 2 replicates these findings on a wider range of negative responses and reveals that the effect of counterfactual thought on willingness to leave the group in the case of a near miss is mediated by attributions of responsibility.

Keywords: Public good dilemmas, counterfactuals, negative affect, exit, near miss effect



Running Head: ON NEAR MISSES IN THE PUBLIC GOOD DILEMMA

On Near Misses in the Public Good Dilemma:

How Upward Counterfactuals Influence Group Stability When the Group Fails

ACCL

Abstract

In two studies we investigated the impact of degree of collective failure in a public good dilemma (near miss vs. large miss) on group members' negative reactions (negative affect, attributions of responsibility for the failure, and intention to leave the group). The results show that upward counterfactual thinking has more impact on members' negative responses when experiencing a near miss rather than a large miss. In Experiment 1, the results show that in the case of a near miss (and not a large miss), negative affect and attributions of responsibility were higher when other-focused counterfactuals rather than self-focused counterfactuals were elicited. Negative affect was found to mediate the effect on attributions of responsibility. Experiment 2 replicates these findings on a wider range of negative responses and reveals that the effect of counterfactual thought on willingness to leave the group in the case of a near miss is mediated by attributions of responsibility.

On Near Misses in the Public Good Dilemma:

How Upward Counterfactuals Influence Group Stability When the Group Fails

In our daily lives, we frequently encounter situations in which our personal interests are at odds with the collective interest (see Komorita & Parks, 1994; Weber, Kopelman, & Messick, 2004). The dilemma of providing public goods represents such a situation of conflict and its essence rests in the circumstance that *provision* of the public good is contingent on the combined contributions of individual group members, whereas the individual's *consumption* of the public good is not related to his/her contribution. Thus, people may enjoy the benefits of the public good even if they do not contribute to its provision. As a result, people may be unwilling to contribute and the collective may fail to provide the public good. How do people deal with failure feedback in public good dilemmas?

Prior research has, for example, revealed that a history of collective failure generally induces individual group members to defect (e.g., Messick & Brewer, 1983; Schroeder, Jensen, Reed, Sullivan, & Schwab, 1983; but see De Cremer & Van Dijk, 2002), and induces people to opt for a structural change by installing a leader or a sanctioning system (e.g., Samuelson, Messick, Rutte, & Wilke, 1984; Samuelson & Messick, 1986). Despite these important insights it is surprising to see that social dilemma research has not focused yet on the psychological impact of collective failure for the individual group members (i.e., how they interpret and emotionally experience group failure). We think it is important to explicitly address such psychological reactions because for the group it directly raises the question of how group stability can be maintained after facing failure (cf. Van Vugt, Jepson, Hart, & De Cremer, 2004). How badly will people feel about having failed as a group? Will they blame others for the failure? Will they blame themselves? And how will these feelings and

interpretations affect group stability, which depends strongly on people's decision to possibly leave their group? Indeed, if too many people exit, the group may in the end have insufficient resources to meet the threshold for providing the public good (e.g. if too many citizens leave then it will become increasingly difficult to collect sufficient contributions to establish public facilities).

To answer these questions, we will draw on insights derived both from the literature on affect, and on the cognitive processes that underlie affective reactions to negative outcomes, in particular on the literature on counterfactual thinking. Specifically, we will (a) focus more on which cognitive processes influence people's responses as a function of negative feedback, (b) examine the emotional consequences of receiving negative feedback information, and (c) investigate how these combined effects may impact on the willingness of group members to leave their group. We will investigate how counterfactual thinking explains people's reactions to failure feedback and whether missing out on the provision of the public good by a small margin (i.e., a near miss) or by a large margin (i.e., a large miss) affects the influence of counterfactual thought.

Counterfactual Thinking and Type of Collective Failure

Recent social dilemma research has revealed that after failure feedback, group members display strong negative affect (assessed by anger and irritation) which undermines group stability by inducing group members to leave the group (see Stouten De Cremer, & Van Dijk, 2005, 2006). To understand the workings of failure feedback in a more comprehensive way we develop an approach in which we identify the *cognitive* strategies that group members use to make sense of failure, and postulate how these cognitive processes consequently influence affect and attributions.

One cognitive process that is of specific relevance to understanding people's reactions to collective failure in social dilemmas is counterfactual thinking.

Counterfactual thinking refers to the idea that people can cognitively simulate how a present situation could have been different (Roese, 1997). Counterfactual thinking is strongly related to causal reasoning (e.g., Kahneman & Varey, 1990; Spellman, Kincannon, & Stose, 2005), and thus to attributions of responsibility.

Interestingly, research has shown that if people fail in achieving a goal, or miss out on an opportunity to obtain positive and favorable outcomes, they often engage in upward counterfactual thinking. That is, they mentally simulate under which conditions the outcome could have been better (Markman, Gavanski, Sherman, & McMullen, 1993; Roese & Olson, 1995; Sanna & Turley, 1996). Importantly, this type of counterfactual thinking elicits negative affect (Markman et al., 1993; Roese, 1997; Roese & Olson, 1995), which subsequently evokes attributions of responsibility (i.e., who is to be blamed for the failure and why I feel so bad; Wells & Gavanski, 1989). Judgments of blame thus arise as a function of immediate negative affect (cf. Haidt, 2001).

A first important assumption that we make in the present paper is that this negative affect (and resulting attributions of responsibility) will particularly be outward focused; i.e., directed towards the others. Research has shown that people are strongly motivated to think positively about themselves and create self-serving judgments and evaluations. This tendency makes that people attribute failures to external factors and successes to their own actions (Baumeister, 1998; Roese & Olson, 2007; Weiner, 1985; Wong & Weiner, 1981). Based on these insights, we postulate that negative feelings should lead particularly to a kind of upward counterfactual thinking that "allows" people to be self-serving. In other words, they

will be particularly motivated to mentally simulate the failure situation in such a way that it is easier to blame others for the failure rather than themselves.

It is important to note, however, that self-serving evaluations are not always that easy to make (despite the general inclination to think in self-serving ways, Roesch & Olson, 2007). Therefore, we argue that negative affect will lead to evaluations that others in the group (rather than oneself) are to be blamed, but particularly so when the type of upward counterfactual facilitates self-serving attributions. The literature to date (at least to our knowledge) has not examined whether upward counterfactuals may differ in how they facilitate versus inhibit self-serving evaluations. In the present research, we introduce a distinction between self-focused counterfactuals versus other-focused counterfactuals. In the case of *other-focused* counterfactuals, group members may mentally revise the existing situation by thinking what the others in the group (and not they) should have done to achieve a better outcome. In the case of *self-focused* counterfactuals, group members engage in counterfactual thinking that includes thinking about how they themselves (and not the others in the group) should have acted to achieve a better outcome.

We expect that other-focused counterfactuals will enhance attributions of responsibility to others, increase negative affect towards these others and will lead to a greater willingness to leave the group. When a self-focus counterfactual is salient, however, it is more difficult to make self-serving judgments and people will be less likely to derogate others by experiencing less negative affect towards those others, attribute less responsibility to them and be less motivated to leave the group.

A second important assumption that we make is that upward counterfactual thinking is more likely to occur when one fails with a small relative to a large margin (Miller & Turnbull, 1990; Kahneman & Miller, 1986). For example, Kahneman and

Tversky (1982) showed that people are more likely to engage in counterfactual thinking when they miss their plane by narrow rather than wide margins, so that they feel worse when having missed the plane by 5 minutes rather than by 30 minutes. Thus, upward counterfactual thinking should emerge particularly when the group just misses out on achieving the threshold.

This idea of “near miss” is very relevant to understand the exact impact of our distinction between self-focused and other-focused counterfactuals. In particular, we argue that when a specific type of counterfactual is activated (i.e., other-focused) affective reactions and willingness to leave may be more intense when the contributions fall just below the threshold for provision than when they fall well below the threshold. If so, this would mean that affective reactions and behavioral intentions may not always become more negative the lower the contributions are. Specifically, we argue that in the case of a near miss group members should be more likely to evaluate who is responsible (Brockner & Wiesenfeld, 1996; Wong & Weiner, 1981). As a consequence, we expected the predicted difference between self-focused counterfactuals and other-focused counterfactuals to be more pronounced in the case of a near miss than in the case of a large miss. Moreover, in the case of a large miss, people will more easily reason that all group members bear some responsibility (i.e., diffusion of responsibility) relative to the situation of a near miss where people will be more motivated to find out whose fault it was. Under the latter conditions, the counterfactual manipulation should thus have a stronger effect on group members’ responses.

In sum, in the present research, we focus on the interactive effect between type of failure feedback and counterfactual thinking by arguing that the type of counterfactual thought (i.e., other-focus counterfactual versus self-focus

counterfactual) will exert a stronger influence on group members' responses (i.e., negative affect, attributions, and willingness to leave the group after a near miss than after a large miss.

Experiment 1

In Experiment 1, we test the prediction that in the case of a near miss negative affect (i.e., anger and irritation) and attributions of responsibility to the other group members will be higher when other-focused counterfactuals are activated relative to self-focused counterfactuals. In the case of a large miss, less difference in negative affect and attributions of responsibility is expected between both counterfactual thought conditions.

Our main dependent measures in Experiment 1 are negative affect and attributions of responsibility. We assess negative affect by using the most representative anger-related emotions used in social dilemma research: irritation and anger. Irritation and anger are strongly related and both emotions are included in the category "anger" (which reflects negative feelings toward others; Russell & Fehr, 1994). The strong connection was also demonstrated in recent social dilemma research that showed that violations of social and fairness norms lead group members to experience anger-related emotions as assessed by a combined score of anger and irritation (Stouten et al., 2006). Irritation and anger have been described as emotions that are elicited in reaction to what others have done, or, in other words, as outward-focused emotions (see Barclay et al., 2005). In light of our prediction that self-serving counterfactuals should facilitate the emergence of negative reactions in the case of a near miss, these specific emotions thus seem particularly relevant.

With respect to our second outcome variable, attributions of responsibility, we assess the extent to which individual group members evaluate the others in the group

as responsible for the failure. Building upon the idea that negative affect motivates causal evaluations (Wells & Gavanski, 1989; see also research showing that anger-related emotions are related to the emergence of attributions; Smith & Ellsworth, 1985; Smith & Lazarus, 1993; Weiner, Graham, & Chandler, 1982), we argue that negative affect shapes the evaluations of who is considered responsible for the collective failure. Therefore, it follows that the interactive effect between failure feedback and counterfactual thought on attributions of responsibility will be mediated by negative affect in such a way that affect will underlie the effect of counterfactual thought on attributions of responsibility in the near miss condition and less so in the large miss condition (i.e., moderated mediation; Preacher, Rucker, & Hayes, 2007).

Method

Participants and design. Eighty-one undergraduate students (33 men and 48 women, average age = 20.87 years, $SD = 2.57$) participated voluntarily and were each paid 7 Euros. They were randomly assigned to a 2 (Failure feedback: Near miss vs. Large miss) x 2 (Counterfactual: Self-focused vs. Other-focused) between-subjects factorial design.

Experimental procedure. After participants arrived at the laboratory they were seated in separate experimental cubicles. They were not able to see one another and each participant's cubicle contained a computer, a table, and a chair. All instructions were given via the computer. At the start of the instructions, participants were informed that they were part of a four-person group and that a number would identify each group member: 1, 2, 3, or 4. In reality, all participants were assigned the number 3.

Introduction to the public good dilemma. Participants were first told about the nature of the task. They were told that each of the four group members possessed an

endowment of 100 chips. Each chip was said to be worth 5 Euro cents.. They were informed that the group as a whole had to contribute a certain number of chips in order to obtain the public good (i.e., the threshold). No exact threshold was specified. Participants were told that the threshold was somewhere between 180 and 300 chips. Some degree of uncertainty with respect to the threshold was introduced to prevent participants to chose the simple rule of equality (which would make it hard to believe that the group would fail with a large margin) and to increase the probability that participants could also assign some blame to their own actions (and not only to the others). Participants were told that if contributions met or surpassed the threshold that the total amount contributed by the group would be multiplied by two and then divided equally amongst all group members, regardless of their contribution. The amount one decided not to contribute would accrue totally to oneself. However, if the group would not succeed in surpassing the threshold, then all the money contributed would be lost. The participants learned that the total earnings would be paid at the end of the study.

Then, participants decided how many chips they wanted to contribute. After this, the failure feedback manipulation was introduced. In the near miss condition, participants were told that the group was 4 chips short of surpassing the threshold. In the large miss condition, participants were told that the group was 52 chips short of surpassing the threshold.

Thereafter, the counterfactual manipulation was introduced. In both conditions, participants finished a sentence and thought about it for a while (they were asked to write down their thoughts on a piece of paper that was lying next to the computer). In the self-focused condition, participants finished the following sentence: The group would have been able to get a better outcome if *you* [italics added]..... In the other-focused

condition, participants finished the following sentence: The group would have been able to get a better outcome if the *others* [italics added]

Then, the dependent measures of Experiment 1 were solicited. All measures were responded to on a 7-point scale (1 = *not at all*, 7 = *very much so*). First, to check the effectiveness of the failure feedback manipulation, participants were asked to indicate the extent in which the group just missed out on achieving the public good. Thereafter, participants' negative affect was assessed by asking them to what extent they felt "angry", and "irritated". These items were combined to form one average negative emotion score ($r = .81, p < .001$). Subsequently, participants were asked to what extent they perceived the others to be responsible for the failure.

Results

Manipulation checks. To check the validity of the counterfactual manipulation, we conducted an inter-rater reliability analysis. Two judges independently coded whether the participants (depending on the experimental condition that the participants were allocated to) either completed the sentence describing how they themselves could have improved the outcome or how the other groups members could have achieved a better outcome. The judges agreed that all participants correctly followed the instructions.

A 2 x 2 ANOVA on the failure feedback manipulation check revealed only a significant effect of Failure feedback, $F(1, 77) = 38.67, p < .001, \eta^2 = .33$, showing that participants in the near miss condition indicated that they just missed out on providing the public good relative to those in the large miss condition ($M_s = 5.57$ vs. 2.72, $SD_s = 2.19$ and 1.87; respectively).

Contribution. On average, participants contributed 53.20 chips ($SD = 20.31$). A 2×2 ANOVA revealed no significant effects of Failure feedback, $F(1, 77) < 1$, of Counterfactual, $F(1, 77) < 1$, or of the Interaction, $F(1, 77) < 1$.¹

Negative affect. A 2×2 ANOVA on the negative affect score revealed, first of all, a significant main effect of Counterfactual, $F(1, 77) = 4.13, p < .05, \eta^2 = .05$: Participants in the other-focused counterfactual condition reported higher negative affect than those in the self-focused counterfactual condition ($Ms = 4.26$ vs. 3.56 , $SDs = 1.70$ and 1.51 ; respectively). Also, a significant interaction emerged, $F(1, 77) = 9.57, p < .005, \eta^2 = .11$ (see Table 1 for means). As predicted, the difference between participants in the self-focused counterfactual condition versus the other-focused counterfactual condition was significant in the near miss condition, $F(1, 77) = 12.99, p < .001$, but not in the large miss condition, $F(1, 77) < 1, p > .45$.

Attributions of responsibility. A 2×2 ANOVA on participants' evaluation of how responsible they regarded the others in the group for the failure revealed only a significant interaction effect, $F(1, 77) = 6.59, p < .05, \eta^2 = .08$ (see Table 1 for means). As predicted, the difference between participants in the self-focused counterfactual condition versus the other-focused counterfactual condition was significant in the near miss condition, $F(1, 77) = 6.05, p < .05$, but not in the large miss condition, $F(1, 77) = 1.35, p > .24$.

Mediation analysis. To test whether negative affect mediated the interactive effect of Failure feedback and Counterfactual on attributions of responsibility of failure, we conducted an additional regression analysis in which we followed the recommendations of Baron and Kenny (1986). We included the previously described independent variables and the responsibility score as dependent variable. Negative affect was added as an additional predictor in the final step of the regression. As we

have shown that our independent variables affected the covariate negative affect and the dependent measure attributions (see our earlier analyses) we further demonstrated that negative affect in the regression significantly affected attributions of responsibility for the failure ($\beta = .70, t = 7.41, p < .001$). Moreover, entering negative affect rendered the interactive effect of Failure feedback and Counterfactual on attributions of responsibility non-significant ($\beta = .41, t = .69, p < .50$).

As a test of mediation, researchers (e.g., MacKinnon et al., 2002) have recommended to directly test the significance of the mediated effect. We used a bootstrap method, developed by Preacher et al. (2007), which allows for a formal test of *moderated* mediation. Specifically, we estimated whether negative affect mediated the effect of Counterfactual on attributions of responsibility in the near miss condition but not in the large miss condition. In accordance with our predictions, Failure feedback was treated as a moderator of the relation between Counterfactual and negative affect. Our effect estimates are based on 5000 bootstrap samples.

The moderated mediation analysis clearly supported our hypothesis: the indirect effect of Counterfactual on attributions of responsibility, via negative affect, was significant in the near miss condition ($z = 3.11, p < .005$), but not in the large miss condition ($z = -.67, p < .50$). Importantly, the bootstrap results for this indirect effect showed that this Beta estimate was significantly larger than 0, which again supports our mediation hypothesis.

Experiment 2

As expected, the results of Experiment 1 showed that our participants experienced stronger negative affect and perceived the others in the group as more responsible for the collective failure when other-focus counterfactuals were made accessible relative to self-focus counterfactuals. Importantly, this pattern of findings

only emerged in the case of a near miss. Further, the moderated mediation analysis indicated that negative affect mediated (at least partly) the effect of counterfactual thought on attributions of responsibility for failure in the near miss condition.

This study is the first to show that the magnitude of failure feedback matters significantly in affecting members' affect and attributions of responsibility as a function of the type of counterfactual that is activated. In Experiment 2 we wanted to replicate these effects to enhance the validity of our novel findings and to introduce some methodological improvements.

In Experiment 1, we assessed negative affect in general, but it is fair to note that this general affect measure does not allow us to draw any inferences about whether one felt negatively toward oneself or toward the others. This distinction can be important for a variety of reasons. First, the mediation analysis reported in Experiment 1 suggests that general negative affect accounted for participants' judgment that the others in the group were responsible for the failure situation. Although this result suggests that the general negative affect score most likely served as a proxy for negative affect toward the others, we need to test whether general negative affect and negative affect towards others yield the same pattern of results. Second, making a distinction between negative affect towards self and the others also allows us to test in a more direct way whether participants engage in a self-serving strategy, and thus would blame the others more for the failure than themselves. Therefore, in Experiment 2, in addition to assessing negative affect in general (as in Experiment 1), we assessed both negative affect toward oneself and toward the others.

Another modification is that in Experiment 2 we also assessed participants' willingness to leave the group or not. Whereas negative affect and blaming others in the group (our other dependent measures) may already be considered as a threat to

group stability, the main threat to stability arises when group can actually leave the group (Van Vugt et al., 2004). Receiving the option to leave the group or not is important in light of the effects that upward counterfactuals may have on group stability. In fact, a recent study by Parks, Sanna and Posey (2003) revealed that upward counterfactuals in a social dilemma might be beneficial as they promote future cooperation. In the present paper we maintain, however, that when it comes to group stability, the effects of upward counterfactuals may not be that positive. Specifically, we posit that upward counterfactuals may increase the willingness to leave the group. Therefore, we predict that individual group members may be more willing to leave the group when an other-focused counterfactual is accessible rather than a self-focused counterfactual, but particularly so in case of a near miss.

Finally, in Experiment 2, we also explore whether negative affect in general, negative affect toward self and others, and attributions of responsibility underlie the predicted interactive effect between failure feedback and counterfactual thinking on willingness to leave the group. First of all, following insights that affect and their subsequent evaluations (i.e., attribution of responsibility in the present research) lead to action tendencies (Ellsworth & Scherer, 2003), we suggest that once people have made sense of the situation and attributed responsibility of failure, they decide on their actions. Second, based on our findings of Experiment 1 that participants' attributions were caused by their negative affect, we expect that primarily attributions of responsibility (which should be shaped first by negative affect) will account for the interactive effect between counterfactual thought and failure feedback on willingness to leave the group. The impact of the different proposed mediators will be tested by employing a multiple mediation analysis (Preacher & Hayes, 2008).

Method

Participants and design. One hundred and twenty-three undergraduate students (32 men and 91 women, average age = 20.39 years, $SD = 1.74$) participated voluntarily and were each paid 7 Euros. They were randomly assigned to a 2 (Failure feedback: near miss vs. large miss) x 2 (Counterfactual: Self-focused vs. Other-focused) between-subjects factorial design.

Experimental procedure. Participants arrived at the laboratory and the procedure was identical to the one used in Experiment 1. At the end of the experiment, the dependent measures were solicited and they were answered on a 7-point scale (1 = *not at all*, 7 = *very much so*). To check for the effectiveness of the failure feedback manipulation the same question as in Experiment 1 was used. Further, participants' negative affect (i.e., anger and irritation) was assessed. First, anger and irritation in general ($r = .61, p < .001$) were assessed. Second, participants were asked how angry and irritated they felt with themselves ($r = .58, p < .001$). Third, participants were asked how angry and irritated they felt towards the others ($r = .60, p < .001$). Thereafter, the same responsibility question as in Experiment 1 was asked. Finally, participants were asked to what extent they wanted to leave the group.

Results

Manipulation checks. To check for the validity of the counterfactual manipulation, we conducted an inter-rater reliability analysis. As in Experiment 2, two judges independently coded whether the participants either completed the sentence describing how they themselves could have improved the outcome or how the other group members could have achieved a better outcome. The judges agreed that all participants correctly followed the instructions.

A 2 x 2 ANOVA on the failure feedback manipulation check revealed only a significant effect of Failure feedback, $F(1, 119) = 81.61, p < .001, \eta^2 = .14$, showing

that participants in the near miss condition indicated that they just missed out on achieving the public good relative to those in the large miss condition ($M_s = 5.96$ vs. 2.75 , $SDs = 1.92$ and 2.03 ; respectively).

Contributions. On average, participants contributed 50.00 chips ($SD = 39.21$). A 2×2 ANOVA revealed no significant effect of Failure feedback, $F(1, 119) < 1$, $p > .66$, of Counterfactual, $F(1, 119) < 1$, $p > .87$, or of the Interaction, $F(1, 119) = 1.13$, $p > .28$.²

Negative affect. A 2×2 ANOVA on the affect score revealed a significant interaction effect, $F(1, 119) = 5.40$, $p < .05$, $\eta^2 = .04$ (see Table 2 for means). As predicted, the difference between participants in the self-focused counterfactual condition versus the other-focused counterfactual condition was significant in the near miss condition, $F(1, 119) = 5.48$, $p < .05$, whereas this was not the case in the large miss condition, $F(1, 119) < 1$, $p > .36$.

Negative affect towards self. A 2×2 ANOVA on the affect score toward oneself revealed only a significant interaction effect, $F(1, 119) = 7.40$, $p < .01$, $\eta^2 = .06$ (see Table 2 for means). As predicted, the difference between participants in the self-focused counterfactual condition versus the other-focused counterfactual condition was significant in the near miss condition, $F(1, 119) = 6.08$, $p < .05$, but not in the large miss condition, $F(1, 119) = 1.85$, $p > .17$.

Negative affect towards the others. A 2×2 ANOVA on the affect score toward others revealed, first of all, a significant main effect of Failure feedback, $F(1, 119) = 5.72$, $p < .05$, $\eta^2 = .05$: Participants in the large miss condition reported higher negative affect with the others than those in the near miss condition ($M_s = 4.11$ vs. 3.47 , $SDs = 1.52$ and 1.51 ; respectively). Also, a significant interaction emerged, $F(1, 119) = 6.16$, $p < .05$, $\eta^2 = .05$ (see Table 2 for means). As predicted, the difference

between participants in the self-focused counterfactual condition versus the other-focused counterfactual condition was significant in the near miss condition, $F(1, 119) = 6.23, p < .05$, but not in the large miss condition, $F(1, 119) < 1, p > .32$.

To examine in greater detail whether negative affect was more strongly expressed towards the others than towards oneself (i.e., representing a self-serving tendency that participants blamed the others more for the group failure), we created a new dependent measure by subtracting the negative affect score towards oneself from the negative affect score towards the others. A positive sign would indicate a more self-serving display in a way that after failure more negative affect is displayed toward the others rather than to oneself, whereas a negative sign would indicate the reversed effect. In line with our prior reasoning, we thus expected that more self-serving displays (i.e. blaming others more) would be found when an other-focused counterfactual was induced relative to a self-focused counterfactual, but particularly so after a near miss.

A 2 x 2 ANOVA on this difference score revealed, first of all, a significant main effect of Failure, $F(1, 119) = 5.53, p < .05, \eta^2 = .04$: Participants in the large miss condition reported stronger negative affect than those in the near miss condition ($M_s = 1.96$ vs. 1.20 , $SD_s = 1.88$ and 1.94 ; respectively). The positive sign thus indicated that a self-serving tendency was present, but particularly so when the group failed by a large margin. In line with our predictions, we also found a significant interaction, $F(1, 119) = 16.18, p < .001, \eta^2 = .12$ (see Table 3 for means). Interestingly, simple effects showed that in the near miss condition participants in the other-focused counterfactual condition showed a stronger self-serving tendency than those in the self-focused counterfactual condition, $F(1, 119) = 14.81, p < .001$, whereas this difference in the large miss condition was marginally significant, $F(1, 119) = 3.24, p < .08$.

Other's responsibility for failure. A 2 x 2 ANOVA on the responsibility score revealed, first of all, a significant main effect of Failure feedback, $F(1, 119) = 12.63$, $p < .01$, $\eta^2 = .10$: Participants in the large miss condition felt more that the others were responsible for the failure than those in the near miss condition ($M_s = 5.64$ vs. 4.76 , $SD_s = 1.17$ and 1.59 ; respectively). Also, a significant interaction emerged, $F(1, 119) = 5.29$, $p < .05$, $\eta^2 = .04$ (see Table 2 for means). As predicted, the difference between participants in the self-focused counterfactual condition versus the other-focused counterfactual condition was significant in the near miss condition, $F(1, 119) = 6.06$, $p < .05$, but not in the large miss condition, $F(1, 119) < 1$, $p < .45$.

Willingness to leave. A 2 x 2 ANOVA on the willingness to leave score revealed, first of all, a significant main effect of Feedback, $F(1, 119) = 6.30$, $p < .05$, $\eta^2 = .05$: Participants in the large miss condition wanted to leave more than those in the near miss condition ($M_s = 4.75$ vs. 4.05 , $SD_s = 1.33$ and 1.79 ; respectively). Also, a significant interaction emerged, $F(1, 119) = 6.41$, $p < .05$, $\eta^2 = .05$ (see Table 4 for means). As predicted, the difference between participants in the self-focused counterfactual condition versus the other-focused counterfactual condition was significant in the near miss condition, $F(1, 119) = 6.04$, $p < .05$, but not in the large miss condition, $F(1, 119) = 1.20$, $p < .28$.

Exploratory analysis: Mediation. We also wanted to explore which variable actually mediates the Failure feedback x Counterfactual interaction on willingness to leave. For this purpose we had to consider several potential mediators. Although the Baron and Kenny approach appears useful (i.e., enter all proposed mediators in the final step in the regression), this approach is subject to a number of important limitations when it comes to assessing the importance of multiple mediators, leading

Preacher and Hayes (2008, p. 882) to note that “We regard the causal steps approach of little utility in the multiple mediator context.”

The use of multiple mediators may lead to problems of interpretation if different mediators overlap in content (West & Aiken, 1997). It has, however, important advantages, relative to the simple mediator approach. First, it is possible to determine to what extent a specific variable mediates the failure feedback x counterfactual thought interaction on willingness to leave, conditional on the presence of other possible mediators in the model. Second, it decreases the risk of biased parameter estimates, as would result if multiple mediation hypotheses were tested with a set of simple mediator models (Judd & Kenny, 1981). Third, and, of central importance to the present study, including several mediators in one model allows us to assess which mediator is most impactful.

We conducted a multiple mediation analysis, using a bootstrap approach taken from Preacher and Hayes (2008) including the previously described independent variables and the willingness to leave score as dependent variable. Our effect estimates were based on 5000 bootstrap samples. This analysis showed that only attributions of responsibility for failure in the regression significantly affected willingness to leave ($\beta = -.37, t = -3.79, p < .0001$), but not negative affect in general ($\beta = -.12, t = -1.29, p < .20$), negative affect towards others ($\beta = .24, t = 1.71, p < .10$), or negative affect towards self ($\beta = .10, t = .74, p < .47$).

Having established by means of multiple mediation testing that only attributions of responsibility for failure significantly mediates the failure feedback x counterfactual thought interaction on willingness to leave, we proceeded by formally testing moderated mediation by means of the same bootstrap method that was employed in Experiment 1 (Preacher et al., 2007). We estimated whether attributions

of responsibility for failure mediated the effect of counterfactual thought on willingness to leave in the near miss condition but not in the large miss condition. In accordance with our predictions, failure feedback was treated as a moderator of the relation between counterfactual thought and negative affect. Our effect estimates are based on 5000 bootstrap samples.

The moderated mediation analysis clearly supported our hypothesis: the indirect effect of counterfactual thought on willingness to leave the group, via attributions of responsibility for failure, was significant in the near miss condition ($z = 1.96, p = .05$), but not in the large miss condition ($z = -.87, p < .39$). Importantly, the bootstrap results for this indirect effect showed that this Beta estimate is significantly larger than 0, which again supports our mediation hypothesis.

General Discussion

Taken together, the present studies consistently show that the type of counterfactual thinking significantly influences members' affect, attributions and decision-making responses, but only so when the group fails to further the collective interest by a small margin. Prior social dilemma research did not devote much attention to the impact of collective feedback on how group members make sense of a failure situation and how this sense-making process influences their immediate reactions. Using insights from the counterfactual literature we explored the idea that the margin by which the group fails in furthering the collective interest matters in how people make sense of the situation at hand and influences their subsequent responses.

Summary of Findings and its Implications

Overall, across two experiments, our results showed that in the case of a near miss our counterfactual manipulation had a significant impact on people's negative

affect, attributions of responsibility and willingness to leave the group, whereas this was not the case after a large miss.

The type of upward counterfactual. To investigate the influence of upward counterfactual thought in a public good setting, we designed a new type of counterfactual manipulation. We employed a manipulation that either facilitated or inhibited people's tendency to engage in self-favoring assessments and responses. Participants either engaged in counterfactual thought reflecting upon how the outcome could have been different when the others in the group (other-focused counterfactual) or when they themselves (self-focused counterfactual) would have acted differently. This difference in reference of the counterfactual (self vs. other) shaped people's negative affect and attributions in different ways.

In Experiment 1, under conditions of a near miss, we found that general negative affect was stronger when other-focused counterfactual thinking rather than self-focused counterfactual thinking was elicited. This effect was replicated in Experiment 2 and, moreover, the same pattern of results was found when the negative affect one felt toward the others (rather than the general negative affect measure) was the dependent variable. These findings imply that, in the case of a near miss, negative affective responses were influenced by our counterfactual manipulation in such a way that they were self-serving. In fact, the overall tendency seemed to be that participants in our studies particularly blamed the others for the near miss rather than themselves. Indeed, the negative affect score was higher in the other-focused than in the self-focused counterfactual conditions. Also, additional analyses in Experiment 2 using the difference score between negative affect toward others and oneself demonstrated that negative affect was always displayed in a stronger way towards the others than

towards oneself, but particularly so when in the case of a near miss an other-focused counterfactual relative to a self-focused counterfactual was activated.

We did not ask the participants what they expected the threshold to be. Note, however, that if participants took the midpoint of the range of 180-300 chips (240 chips) as a reasonable estimate, participants in the large miss conditions could have concluded that their group had contributed 188 chips (being 52 chips short). With mean contributions of the participants being 53 chips (Study 1) and 50 chips (Study 2), this would imply that the participants would have contributed *more* than the others; On average, the three fellow group members would have contributed only 45 chips in Study 1 ($[188-53]/3$) and 46 chips in Study 2 ($[188-50]/3$). In contrast, in the near miss conditions, the conclusion would then be that they had contributed *less* than the others: On average, the three fellow group members would have contributed 61 chips in Study 1 ($[236-53]/3$) and 62 chips in Study 2 ($[236-50]/3$). If these are reasonable assumptions, this would make our data even more impressive in the sense that we observed that, especially in the other-focused conditions, participants blamed themselves less after a near miss than after a large miss.

On a related note, the pervasiveness of this self-serving attribution strategy may lead one to wonder whether the refusal to attribute failure to one's own behavior goes hand in hand with a tendency to attribute the failure of the group to structural features such as the difficulty of the threshold that needed to be surpassed. In the current settings we introduced some uncertainty in the provision threshold to allow for internal attributions of blame. But of course, this uncertainty could also allow for external attributions to the threshold level which could further facilitate self-serving attributions. It might thus be that under conditions of a near miss and self-focused counterfactuals people have higher estimations of the threshold to justify to

themselves that they did not do more in terms of contributions (see also Wit & Wilke, 1998).

The magnitude of collective failure. Our findings demonstrated that the effect of counterfactual thought was moderated by the type of collective feedback offered. These results thus point out that to understand the experience of collective failure feedback it is important not only to focus on the valence of the feedback (i.e., “did we succeed or not?”), but even more so on the magnitude of failure (near vs. large miss). Indeed, our results provided evidence for an intriguing observation that a large miss is not necessarily worse than a near miss. In fact, depending on the margin of collective failure, group members’ upward counterfactual thinking influenced (near miss) or did not influence (large miss) their negative reactions. As we pointed out earlier, in the case of a large miss all group members will be perceived as somewhat responsible (i.e. diffusion of responsibility) whereas in the case of a near miss group members will be less likely to take some part of the blame and be more motivated to search for the one member responsible, which allows counterfactuals to shape the blaming process.

What drives what? Our findings also provided some interesting insights with respect to how people arrive at attributions of failure and decisions to leave when a near miss situation is encountered. One common observation in the literature is that little is known about the underlying processes that shape people’s responses when the group fails in furthering the collective interest. Experiment 1, first of all, showed that, negative affect shaped attribution judgments. In the near miss condition, participants felt more irritated and angry when other-focused counterfactuals rather than self-focused ones were activated, and this led people to attribute more responsibility of failure to the others in the group. Experiment 2 further showed that attributions of

responsibility explained participants' willingness to leave in the case of a near miss. Thus, in the case of a near miss, stronger negative affect is experienced when other-focused rather than self-focused counterfactuals are activated and these affective responses drive attribution judgments, which, in turn drive decision-making.

In this context, it is interesting to note that Experiment 2 showed that - in the case of a near miss - participants were less willing to leave the group when they had generated own-focused counterfactuals. A possible reason for not wanting to leave the group is that one wants to take constructive action to further the group's interests. This interpretation accords with the suggestion of Parks et al. (2003) that upward counterfactuals may be beneficial for the group, although the current findings thus show that this beneficial effect requires a self-focus.

In Conclusion

The current investigation demonstrates the importance of identifying the type of upward counterfactual thought group members can engage in when failing to establish a public good. Specifically, our results show that the impact of these types of counterfactual thought depends on the margin of collective failure - a near miss revealing different effects than a large miss. As such, our results emphasize the importance to focus in greater detail on how people interpret failures to further the collective interest and how the affective and cognitive processes involved set the agenda for future decisions.

References

- Barclay, L. J., Skarlicki, D. P., & Pugh, S. D. (2005). Exploring the role of emotions in injustice perceptions and retaliation. *Journal of Applied Psychology, 90*, 629-643.
- Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed.; pp. 680-740). New York: McGraw-Hill.
- Baron, R. M., & Kenny, D. A. (1986). The moderator mediator variable distinction in social-psychological research – conceptual, strategic, and statistical consideration. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Brockner, J., & Wiesenfeld, B. M. (1996). An integrative framework for explaining reactions to decisions: Interactive effects of outcomes and procedures. *Psychological Bulletin, 120*, 189-208.
- De Cremer, D., & Van Dijk, E. (2002). Reactions to group success and failure as a function of identification level: A test of the goal-transformation hypothesis in social dilemmas. *Journal of Experimental Social Psychology, 38*, 435-442.
- Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572-595). New York: Oxford University Press.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Bulletin, 108*, 814-834.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*, 602-619.
- Kahneman, D., & Miller, D.T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136-153.
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P.

Slovic, & A. Tversky (Eds.), *Judgment under uncertainty* (pp. 201-218).

Cambridge, England: Cambridge University Press.

Kahneman, D., & Varey, C. A. (1990). Propensities and counterfactuals: The loser that almost won. *Journal of Personality and Social Psychology*, 59, 1101-1110.

Komorita, S. S., & Parks, C. D. (1994). *Social dilemmas*. Dubuque, IA: Brown & Benchmark.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variables effects. *Psychological Methods*, 7, 83-104.

Markman, K. D., Gavanski, L., Sherman, S. J., & McMullen, M. N. (1993). The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology*, 29, 87-109.

Messick, D. M., & Brewer, M. B. (1983). Solving social dilemmas. In L. Wheeler, & P.R. Shaver (Eds.), *Review of Personality and Social Psychology* (Vol. 4, pp. 11-44). Beverly Hills: Sage publications.

Miller, D. T., & Turnbull, W. (1990). The counterfactual fallacy: Confusing what might have been with what ought to have been. *Social Justice Research*, 4, 1-19.

Parks, C. D., Sanna, L. J., & Posey, D.C. (2003). \retrospection in social dilemmas: Thinking about the past affects future cooperation. *Journal of Personality and Social Psychology Bulletin*, 84, 988-996.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in simple and multiple mediator models. *Behavior Research Methods*, 40, 879-891.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate*

Behavioral Research, 42, 185-227.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, 121, 133-148.

Roese, N. J., & Olson, J. M. (1995). Counterfactual thinking: An overview. In N. J. Roese and J. M. Olson (Eds.), *What might have been: the social psychology of counterfactual thinking* (pp. 1-55). Mahwah, NJ: Erlbaum.

Roese, N. J., & Olson, J. M. (2007). Better, stronger, faster: Self-serving judgments, affect regulation, and the optimal vigilance hypothesis. *Perspectives on Psychological Science*, 2, 124-141.

Russell, J. A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: varieties of anger. *Journal of Personality and Social Psychology*, 67, 186-205.

Samuelson, C. D., Messick, D.M. (1986). Inequities in access to and use of shared resources in social dilemmas. *Journal of Personality and Social Psychology*, 51, 298-324.

Samuelson, C.D., Messick, D.M., Rutte, C., & Wilke, H. (1984). Individual and structural solutions to resource dilemmas in two cultures. *Journal of Personality and Social Psychology*, 47, 94-104.

Sanna, L. J., & Turley, K. J. (1996). Antecedents to spontaneous counterfactual thinking: effects of expectancy violation and outcome valence. *Personality and Social Psychology Bulletin*, 22, 906-919.

Schroeder, D. A., Jensen, T. D., Reed, A. J., Sullivan, D. K., & Schwab, M. (1983). The actions of others as determinants of behavior in social dilemmas. *Journal of Experimental Social Psychology*, 19, 522-539.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813-838.

Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes,

and the emotions. *Cognition and Emotion*, 7, 233-269.

- Spellman, B.A., Kincannon, A., & Stose, S.(2005). The relation between counterfactual and causal reasoning. In D. Mandel, D. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking*. London: Routledge.
- Stouten, J., De Cremer, D., & Van Dijk, E. (2005). All is well that ends well, at least for proselves: Emotional reactions to equality violation as a function of social value orientation. *European Journal and Social Psychology*, 35, 767-783.
- Stouten, J., De Cremer, D., & Van Dijk, E. (2006). Violating equality in social dilemmas: Emotional and retributive reactions as a function of trust, attribution, and honesty. *Personality and Social Psychology Bulletin*, 32, 894-906.
- Van Vugt, M., Jepson, S., Hart, C., & De Cremer, D. (2004). Why autocratic leadership might fail in solving public good dilemmas: The importance of group stability. *Journal of Experimental Social Psychology*, 40, 1-12.
- Weber, J. M., Kopelman, S., & Messick, D. M. (2004). A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness. *Personality and Social Psychology Review*, 8, 281-307.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92, 548-573.
- Weiner, B., Graham, S., & Chandler, C. (1982). Pity, anger, and guilt: An attributional analysis. *Personality and Social Psychology Bulletin*, 8, 226-232.
- Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, 56, 161-169.
- West, S. G., & Aiken, L. S. (1997). Toward understanding individual effects in multicomponent prevention programs: Design and analysis strategies. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention:*

Methodological advances from alcohol and Multiple Mediation 44 substance abuse research (pp. 167-209). Washington, DC: American Psychological Association.

Wit, A., & Wilke, H. (1998). Public good provision under environmental and social uncertainty. *European Journal of Social Psychology*, 28, 249-256.

Wong, P. T. P. & Weiner, B. (1981). When people ask “why” questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology*, 40, 650-663.

Footnotes

¹ Including contributions in the first session as a covariate did not change the results.

² Again, including contributions in the first session as a covariate did not change the results.

Table 1

Negative Affect and Attribution of Responsibility for Failure as a Function of Failure

Feedback and Counterfactual (Experiment 1)

Dependent		Failure Feedback	
Variable	Counterfactual	Near Miss	Large Miss
Negative Affect	Self	2.88 (1.31)	4.26 (1.78)
	Others	4.63 (1.26)	3.90 (1.69)
Attribution of responsibility for failure	Self	3.45 (1.84)	4.86 (1.49)
	Others	4.75 (1.44)	4.25 (1.86)

Note. Higher scores indicate higher levels of negative affect and attributions of responsibility. Standard deviations are given in parentheses.

Table 2

Negative Affect, Negative Affect towards Self, Negative Affect towards others and Attribution of Responsibility for Failure as a Function of Failure Feedback and Counterfactual (Experiment 2)

Dependent		Failure Feedback	
Variable	Counterfactual	Near Miss	Large Miss
Negative Affect	Self	3.18 (1.70)	4.32 (1.46)
	Other	4.12 (1.59)	3.97 (1.38)
Negative Affect towards self	Self	2.68 (1.26)	1.92 (1.18)
	Other	1.84 (0.93)	2.36 (1.26)
Negative Affect toward others	Self	2.98 (1.55)	4.29 (1.57)
	Other	3.94 (1.32)	3.92 (1.46)
Attribution of responsibility for failure	Self	4.33 (1.66)	5.77 (1.23)
	Others	5.20 (1.39)	5.51 (1.12)

Note. Higher scores indicate higher levels of negative affect, negative affect towards self, negative affect toward others, and attributions of responsibility for failure.

Standard deviations are given in parentheses.

Table 3

Difference Score Negative Affect as a Function of Failure Feedback and

Counterfactual (Experiment 2)

Dependent		Failure Feedback	
Variable	Counterfactual	Near Miss	Large Miss
Difference score	Self	.30 (1.92)	2.37 (1.86)
	Others	2.10 (1.51)	1.56 (1.84)

Note. Higher scores indicate higher negative affect toward others. Standard deviations are given in parentheses.

Table 4

*Willingness to Leave as a Function of Failure Feedback and Counterfactual**(Experiment 2)*

Dependent Variable		Failure Feedback	
	Counterfactual	Near Miss	Large Miss
Willingness to	Self	3.56 (1.85)	4.96 (1.40)
Leave	Others	4.55 (1.61)	4.55 (1.25)

Note. Higher scores indicate higher levels of willingness to leave. Standard deviations are given in parentheses.