

When contributions make a difference: Explaining order effects in responsibility attribution

Tobias Gerstenberg · David A. Lagnado

Published online: 15 May 2012
© Psychonomic Society, Inc. 2012

Abstract In two experiments, we established an order effect in responsibility attributions. In line with Spellman (Journal of Experimental Psychology: General 126: 323–348, 1997), who proposed that a person's perceived causal contribution varies with the degree to which it changes the probability of the eventual outcome, Experiment 1 showed that in a team challenge in which the players contribute sequentially, the last player's blame or credit is attenuated if the team's result has already been determined prior to her acting. Experiment 2 illustrated that this attenuation effect does not overgeneralize to situations in which the experienced order of events does not map onto the objective order of events; the level of the last person's performance is only discounted if that person knew that the result was already determined. Furthermore, Experiment 1 demonstrated that responsibility attributions remain sensitive to differences in performance, even if the outcome is already determined. We suggest a theoretical extension of Spellman's model, according to which participants' responsibility attributions are determined not only by whether a contribution made a difference in the actual situation, but also by whether it would have made a difference had things turned out somewhat differently.

Keywords Responsibility attribution · Order effect · Counterfactuals · Causal chain · Causal reasoning · Judgment and decision making · Social cognition

Imagine that you are the coach of your country's relay team. In what order would you make your runners compete? Would you put the best runner first, last, or in one of the

middle positions? How responsible would the runners in the different positions be if the team won or lost? In this article, we are interested in people's perceptions of the extent to which individuals carry responsibility for their groups' results in situations in which the group members contribute sequentially.

According to a simple counterfactual analysis, each of the individual events in a causal chain qualifies equally as a cause of the final effect. If any of the events in the chain had not occurred, the effect would also not have occurred. However, several studies have shown systematic differences as to which events in a chain are judged to be more causal (Miller & Gunasegaram, 1990) or more likely to be mentally undone in order to prevent the outcome from happening (Wells, Taylor, & Turtle, 1987).

In an influential article, Spellman (1997) proposed the *crediting causality model* (CCM), which predicts that an event's perceived causal contribution varies with the extent to which it changes the probability of the eventual outcome.¹ The more an event changes the outcome's probability, the more it is judged to be causal. Accordingly, the model predicts a *primacy effect* when the first event in a causal chain changes the probability of the outcome more than any of the later events. Conversely, it predicts a *recency effect* when the probability change is greatest for the final event in a chain.

Since its proposal, several shortcomings of the CCM have been demonstrated. Importantly, because the model

T. Gerstenberg (✉) · D. A. Lagnado
Cognitive, Perceptual and Brain Sciences,
University College London,
26, Bedford Way,
London WC1H 0AP, UK
e-mail: t.gerstenberg@ucl.ac.uk

¹ Although Spellman's (1997) model was originally developed for judgments of *causal* contributions, it was also used to predict judgments about blame and experienced guilt. In our experiments, we used the term *blame* for negative responsibility and *credit* for positive responsibility, to highlight the outcome-dependent valences of responsibility attributions. However, we acknowledge that these terms are not equivalent and that situations exist in which the results will be affected by choice of terminology (see, e.g., Robbenolt, 2000).

predicts causality ratings merely on the basis of the statistical notion of probability change, it is insensitive to the ways in which these changes are brought about. However, studies have shown that voluntary human actions are preferred over physical events as causes (Lagnado & Channon, 2008), even when the changes in probability are identical (Hilton, McClure, & Sutton, 2010). Furthermore, Mandel (2003) showed that a later event can receive a higher causal rating, even though an earlier event has already increased the probability of the outcome almost to certainty. If a victim has been poisoned first but is then killed in a car crash, people select the car crash as the cause of death rather than the poison, despite the fact that the poison had already increased the probability of death to certainty.

In this article, we highlight a different problem that has not been addressed by previous research and motivate a theoretical extension of the model. The CCM predicts that people's attributions of responsibility are determined by comparing what actually happened with what would have happened had an event in *this particular situation* been different. However, we argue that attributions of responsibility are affected not only by the degree to which an event made a difference in the particular situation in which it occurred, but also by what would have happened had the event of interest been different in *other, similar situations* (see Chockler & Halpern, 2004, for a formal model that incorporates this idea, and Gerstenberg & Lagnado, 2010, for empirical support).

Consider the example of a team relay, mentioned earlier. If the performances of the first three runners in a team were very poor, the probability of the team winning before the fourth runner started would be essentially zero and could not be increased any more, irrespective of the fourth runner's performance. Since the last runner's performance did not make a difference to the team outcome in this particular situation, the CCM would predict that this runner's responsibility would be low, independent of whether she performed well or poorly.

However, when we consider not only the actual situation, but also other possible situations, it becomes clear how responsibility attributions could still be sensitive to differences in performance. The athlete who performed well despite the certainty of the team's loss could have made a difference to the outcome *if* the other team members' performance had been better. In contrast, an athlete who performed poorly in the same situation would send an ambiguous signal: It could be that she did not try hard, because the team outcome was already determined. Yet it is also possible that the athlete would not have performed better even if a situation arose in which the final contribution was required. This difference in uncertainty over whether or not an athlete is capable of performing well licenses a differential attribution of responsibility. Despite the fact that the difference in performance did not matter in the actual

situation, it would have made a difference in situations in which the performance of the other three team members had been better.

In the present study, we explored how (1) level of performance and (2) the extent to which a contribution was critical to the result, as measured by the change of the outcome's probability that the contribution induced, affect people's perceptions of how responsible each contribution is for the eventual outcome. In line with CCM, we expected that the extent to which identical performance would be seen as responsible for the team's result would vary depending on whether or not the result was already determined. We also predicted that responsibility attributions would be influenced not only by how much a person's contribution made a difference in the actual situation, but also by whether it could have made a difference in other possible situations. Hence, we expected that the level of performance of an individual player would influence how responsible she was seen to be for the team's outcome, even in situations in which the result had already been determined prior to her performance.

Experiment 1

The participants acted as external observers evaluating the performance of different teams in the Olympic qualifiers of an invented sport. Each of 32 countries was represented by a team of three athletes. The athletes performed their routines individually and received a score from a panel of judges ranging from 0 (*very bad performance*) to 10 (*excellent performance*). The participants were instructed that the average performance in the competition was 5 points. A country would qualify for the Olympics if its team scored 15 or more points in total. It did not matter how many points a team got once they were above this qualifying standard. The participants were informed that the athletes performed their individual routines sequentially and that later athletes knew how their previous teammates had performed. For each of the 32 teams, the participants experienced two different phases. In the *probability rating phase*, they saw the scores of each of the three athletes sequentially and, after each athlete's score, indicated on a slider how likely they thought it was that the team would qualify (see Fig. 1a). The slider ranged from 0 (*definitely not*) to 100 (*definitely yes*) and was initialized at the midpoint. The progress bar at the top of the screen showed how many points were still required for the team to qualify and was updated after each athlete's score. Once the team qualified, the progress bar turned green. If the team could not qualify anymore, the bar turned red.

In the *responsibility attribution phase*, each athlete's score was shown simultaneously in a table (see Fig. 1b). The participants were asked, "To what extent is each of the athletes responsible for their team's success or failure to

Team #2

0 15

Isthmus

Score

| | |
|--------------|---|
| 1st Teammate | 6 |
| 2nd Teammate | 7 |

Do you think the team is going to qualify?

definitely not | definitely yes

Rules OK

a) Probability updating phase

Isthmus

Isthmus qualified!

Score Credit

| | | |
|--------------|---|--|
| 1st Teammate | 6 | |
| 2nd Teammate | 7 | |
| 3rd Teammate | 2 | |

none high

Rules OK

Once you are happy with your answers please press OK to go to the next round

b) Responsibility attribution phase

Fig. 1 Screenshots of the experiment

qualify?” If the team qualified, the participants attributed credit (green sliders ranging from the center to the right). If the team did not qualify, participants attributed blame (red sliders ranging from the center to the left). The sliders for each athlete ranged from 0 (*none*) to 10 (*high*) and could be moved independently; that is, they did not have to sum to a certain value.

Table 1 shows the patterns of scores that were used in the experiment. We systematically varied the scores of the third athlete to be either low or high for different scores by the first two athletes. This approach allowed us to compare how an identical performance of the third athlete would be evaluated, as a function of whether the result was already *certain* prior to the final athlete’s performing or was still *uncertain*. There were two possible ways in which the team’s result could already have been determined by the scores of the first two athletes. A team’s loss was certain if the sum of the first two athletes’ scores was 4 points or less. Because the maximum score that an athlete could achieve in the challenge was 10, it was impossible in this case for the third athlete to allow the team to win. Likewise, a team’s win was certain prior to the third athlete’s performance if the first two athletes’ scores added up to 15 or more points.

A consequence of this design was that, while we kept the *absolute* performance of the third athlete identical, in the

different situations the *relative* performance as compared to the teammates varied. The final athlete performed relatively well in the *certain loss* as compared to the *uncertain loss* cases, and relatively poorly in the *certain win* as compared to the *uncertain win* situations. Because our main interest concerned the effect of the (un)certainty of the outcome on the attributions for the third athlete, we controlled for the effects of relative performance by including eight additional cases in which the scores of all three athletes were identical. Here, all athletes either scored 1 (or 2) in the *certain loss* cases, 3 (or 4) in the *uncertain loss* cases, 6 (or 7) in the *uncertain win* cases, and 8 (or 9) in the *certain win* cases. Any differences between the three athletes in these situations can only be explained in terms of order effects.

The main target of interest in our design was the third athlete. We hypothesized that both her performance and the certainty of the team’s result prior to her turn would affect the participants’ responsibility attributions. In line with the CCM, we predicted that the third athlete would receive less credit for an identical performance if the result was already certain rather than still uncertain. Likewise, we predicted that the athlete would receive less blame if the team had already certainly missed the qualification threshold prior to her turn. However, in contrast to the CCM, we expected the third athlete’s blame for losses to be higher and her credit for wins to be lower when she received a low rather than a high score, even in situations in which the results were already certain.

Method

Participants A group of 41 (22 female, 19 male) participants recruited through the UCL subject pool took part in this experiment. Their mean age was 23.1 years ($SD = 2.5$).

Materials The program was written in Adobe Flash CS5.

Design For the 24 patterns in which the scores of the three athletes were nonidentical, the experiment followed a within-subjects 2 (Result: win vs. loss) \times 2 (Certainty of Outcome: uncertain vs. certain) \times 2 (Performance of the Third Athlete: low vs. high score) design. For the eight patterns in which the scores of the three athletes were identical, the experiment followed a within-subjects 2 (Result) \times 2 (Certainty of Outcome) design.

Procedure The study was carried out online.² After having read the instructions, the participants did one practice trial in which the different components of the screen were explained. A set of four comprehension check questions ensured that the

² The experiment can be accessed here: http://www.ucl.ac.uk/lagnado-lab/experiments/demos/order_demo.html.

Table 1 Patterns of athletes' scores in the experiments

| Situation | Scores | Athlete 1 | Athlete 2 | Athlete 3 low | Athlete 3 high |
|----------------|--------------|-------------|-------------|---------------|----------------|
| Certain loss | Nonidentical | 1 | 1 | 4 | 8 |
| | | 1 | 2 | 3 | 7 |
| | | 3 | 1 | 2 | 6 |
| | <i>Mean</i> | <i>1.67</i> | <i>1.33</i> | <i>3</i> | <i>7</i> |
| | Identical | 1 | 1 | 1 | |
| Uncertain loss | Nonidentical | 2 | 2 | 2 | |
| | | 4 | 2 | 4 | 8 |
| | | 2 | 5 | 3 | 7 |
| | <i>Mean</i> | <i>3.33</i> | <i>3.67</i> | <i>3</i> | <i>7</i> |
| | Identical | 3 | 3 | 3 | |
| Uncertain win | Nonidentical | 4 | 4 | 4 | |
| | | 5 | 6 | 4 | 8 |
| | | 7 | 5 | 3 | 7 |
| | <i>Mean</i> | <i>6</i> | <i>6</i> | <i>3</i> | <i>7</i> |
| | Identical | 6 | 6 | | 6 |
| Certain win | Nonidentical | 7 | 7 | | 7 |
| | | 7 | 9 | 4 | 8 |
| | | 9 | 8 | 3 | 7 |
| | <i>Mean</i> | <i>8.33</i> | <i>8.67</i> | <i>3</i> | <i>7</i> |
| | Identical | 8 | 8 | | 8 |
| | | 9 | 9 | | 9 |

Mean=average score of each athlete for the patterns with nonidentical scores

participants had understood the task. On average, they answered 89 % of the comprehension check questions correctly. After they had answered each of the questions, the correct

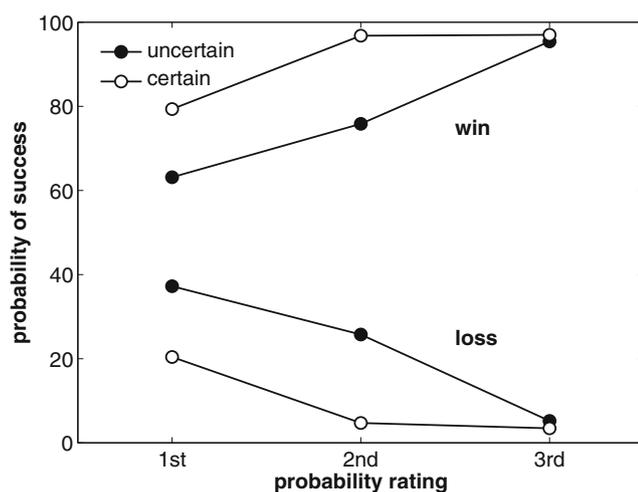


Fig. 2 Mean rated probabilities of success for wins (top) and losses (bottom) as a function of whether the outcome was uncertain (black) or certain (white) after the second player's score. The data are aggregated across Experiments 1 and 2

solution was displayed. The participants then evaluated the performance of 32 teams in the probability rating phase (Fig. 1a) and the attribution phase (Fig. 1b), as described above. If a team did not qualify, the participants attributed blame; otherwise, they attributed credit. Throughout the experiment, they could remind themselves of the rules by clicking on the "Rules" button at the bottom left corner of the screen. The median time that it took participants to finish the study was 18.8 min.

Results

Probability updating phase Figure 2 shows the mean probabilities of success ratings for wins and losses, separated for situations in which the outcome was either already certain after the second athlete's score or still uncertain. The results of the probability updating phase did not differ between Experiments 1 and 2; hence, we report here the aggregated data of both experiments.³

³ Experiments 1 and 2 were identical except for a manipulation in the instructions that did not affect the probability updating phase (see below).

For losses, the participants' probability-of-success ratings after the second athlete's score was revealed were significantly lower in the certain ($M = 4.7, SD = 7.6$) than in the uncertain cases ($M = 25.7, SD = 9.9$), $t(118) = -26.44, p < .05, r = .43$. For wins, the participants' probability ratings were significantly higher in the certain ($M = 96.8, SD = 8$) than in the uncertain ($M = 75.8, SD = 9.85$) cases, $t(118) = -28.18, p < .05, r = .44$.

Responsibility attribution phase First, we wanted to test the extent to which the blame and credit ratings for the third athlete would vary as a function of performance and of whether or not the team's result was already certain after the second athlete's score. Two separate 2 (certainty: certain vs. uncertain) \times 2 (performance: low vs. high score) repeated measures ANOVAs for losses and wins were conducted on the ratings for the third athlete in the nonidentical cases (see Fig. 3a top).

For losses, we found significant main effects of performance, $F(1, 40) = 129.33, p < .05, \eta_p^2 = .764$, and certainty, $F(1, 40) = 6.86, p < .05, \eta_p^2 = .146$, but no interaction. The third athlete was blamed more if she received a low rather than a high score. Furthermore, her blame ratings were lower when the team had already certainly missed the qualification criterion, as compared to when the outcome was still uncertain. Crucially, the effect of performance significantly influenced the athlete's blame ratings for situations in which the outcome was still uncertain, $t(40) = 11.37, p < .05, r = .47$, as well as when the outcome was already determined, $t(40) = 8.36, p < .05, r = .42$.

For wins, we found significant main effects of performance, $F(1, 40) = 66.03, p < .05, \eta_p^2 = .623$, and of certainty, $F(1, 40) = 31.76, p < .05, \eta_p^2 = .443$, but no interaction effect. The third

athlete received more credit for a high than for a low score. Also, credit attributions were higher if the result was still uncertain, as compared to when it was certain. Again, the effect of performance significantly influenced the athlete's credit ratings for situations in which the outcome was still uncertain, $t(40) = -7.76, p < .05, r = .40$, as well as when it was already determined, $t(40) = -6.84, p < .05, r = .38$.

As outlined above, the predicted order effect and the relative performance effect go in the same direction for the cases in which the scores of the three athletes were nonidentical. For example, the third athlete performed relatively well as compared to the others in situations in which the loss was certain rather than uncertain after the second athlete's performance (see Table 1). Hence, we analyzed the situations separately in which all of the athletes had identical scores. Any differences for these cases could only be explained with respect to the order of performance.

Figure 3a (bottom) shows the mean blame ratings for losses and credit ratings for wins attributed to all three athletes in the team in situations in which the result was either uncertain or certain. To evaluate whether the third athlete's ratings varied as a function of certainty of outcome, we compared the difference in the average attributions to the first two athletes with the attributions to the third athlete. For losses, this difference was significantly greater in the certain cases ($M = -1.97, SD = 3.77$) than in the uncertain cases ($M = 0.11, SD = 1.81$). The third athlete received significantly less blame for an identical performance if the result was already certain as compared to uncertain, $t(40) = -3.88, p < .05, r = .30$. For wins, likewise, the third athlete received significantly less credit for an identical performance if the result was already certain ($M = -2.79, SD = 4.18$) than if it was uncertain ($M = -0.71, SD = 2.16$), $t(40) = -3.29, p < .05, r = .28$.

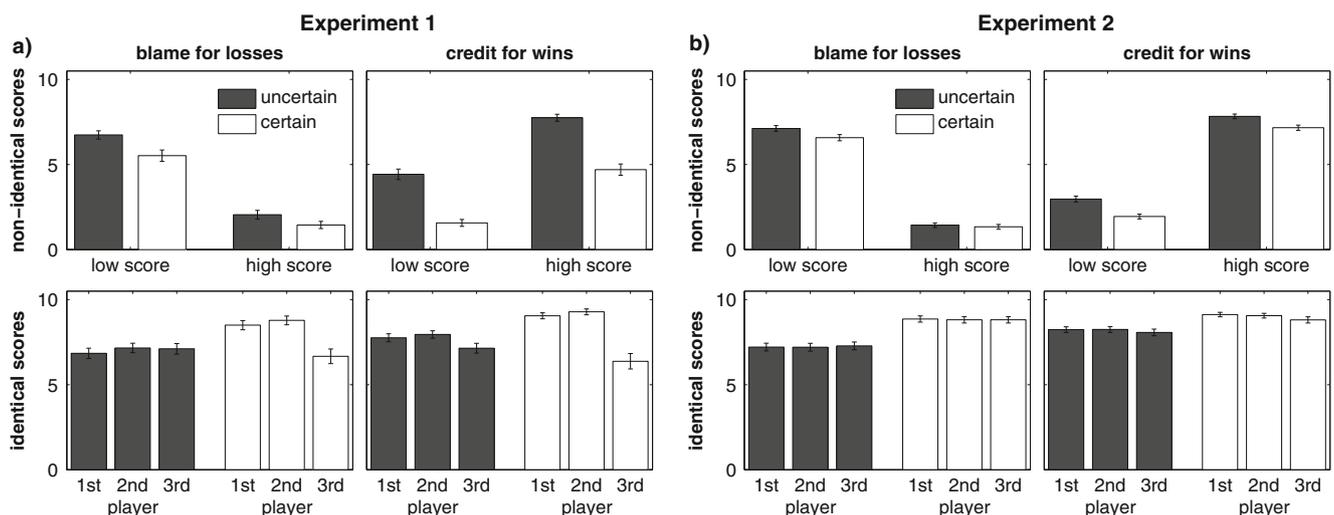


Fig. 3 Mean blame and credit attributions for (a) Experiment 1 and (b) Experiment 2 (collapsed across both conditions). The top panels show mean attributions for the third player in the nonidentical-score cases,

and the bottom panels show mean attributions for all three players in the identical-score cases. Error bars denote $\pm 1 SEM$

Discussion

The results of [Experiment 1](#) revealed an *attenuation effect*: Blame and credit attributions to the last athlete were reduced when the result was certain rather than uncertain. However, how much blame (or credit) an athlete received for the team's loss (or win) also depended to a large extent on her performance. Importantly, even when the result of the team challenge was already determined, an athlete still received more credit and less blame for a good than for a bad performance.

Experiment 2

In [Experiment 2](#), we investigated whether the attenuation effect observed in [Experiment 1](#) for the situations in which the outcome was already determined would overgeneralize to situations in which it would arguably be inappropriate. As outlined above, an athlete who achieves a low score in a situation in which *she knows* that the outcome is already determined sends an ambiguous signal: She could achieve a low score either for not trying hard or for just not being capable of performing well. Crucially, an athlete can only justify not trying hard in a situation in which she knows that the team result is already certain. Our paradigm allowed us to dissociate the “objective certainty” of the team result from the perspective of an external observer from the “subjective certainty” of the team result from the perspective of an athlete in the team. For [Experiment 2](#), we kept the experienced order of events unchanged but altered the knowledge states of the athletes. As in [Experiment 1](#), the participants experienced the scores of the three athletes sequentially; however, they were instructed that the athletes did *not* know about each other's scores.

Spellman (1997) has shown that attributions are influenced not only by the experienced order of events but also by the order in which the events occurred in the world. To test for any effects of the objective order of events, we manipulated whether athletes were described as competing simultaneously (Exp. 2a) or sequentially (Exp. 2b). Would participants show reduced blame attributions for losses and credit attributions for wins when they knew that the result was already determined? Or would they appreciate that the athletes did not know each other's scores, and hence would show no attenuation effect as a function of the certainty of the result?

Method

Participants A group of 56 participants (42 female, 14 male) from the UCL subject pool participated in [Experiment 2a](#) and 22 participants (15 female, 7 male) recruited through Amazon Mechanical Turk took part in [Experiment 2b](#). The mean age was 25.36 years ($SD = 9.8$).

Design and materials The design and materials were identical to those of [Experiment 1](#). Objective Order of Events (simultaneous vs. sequential) was included as between-subjects factor.

Procedure The procedure was identical to that of [Experiment 1](#), except for a minimal change in the instructions: The participants were informed that the athletes of each team performed their individual routines either simultaneously (Exp. 2a) or sequentially (Exp. 2b). Importantly, the participants were instructed that the athletes *did not know* their teammates' scores. On average, the participants answered 89 % of the comprehension questions correctly. The median time that it took participants to complete the study was 16.5 min.

Results

Responsibility attribution phase None of the comparisons were significantly influenced by the between-subjects factor Objective Order of Events. Hence, we combined the data of both conditions.

Again, the blame and credit ratings for the third athlete varied with how the athlete performed and, although to a much lesser extent, with whether or not the result was already certain. Figure 3b (top) shows the blame and credit attributions of the third athlete for the situations in which the scores of the three athletes were nonidentical. For losses, we found significant main effects of performance, $F(1, 77) = 433.08$, $p < .05$, $\eta_p^2 = .849$, and certainty, $F(1, 77) = 13.53$, $p < .05$, $\eta_p^2 = .149$, as well as an interaction effect, $F(1, 77) = 4.52$, $p < .05$, $\eta_p^2 = .55$. For wins, we found significant main effects of performance, $F(1, 77) = 544.32$, $p < .05$, $\eta_p^2 = .876$, and certainty, $F(1, 77) = 24.77$, $p < .05$, $\eta_p^2 = .243$, but no interaction effect.

Importantly, the difference between the blame attributions in the certain cases and the uncertain cases was significantly larger in [Experiment 1](#) ($M = 0.91$, $SD = 2.23$) than in [Experiment 2](#) ($M = 0.23$, $SD = 0.70$), $t(117) = 2.12$, $p < .05$, $d = 0.41$. Likewise, the differences between the credit attributions as a function of outcome certainty were larger in [Experiment 1](#) ($M = 2.95$, $SD = 3.35$) than in [Experiment 2](#) ($M = 0.40$, $SD = 0.81$), $t(117) = 4.73$, $p < .05$, $d = 0.91$.

Figure 3b (bottom) shows the blame and credit ratings of all three athletes for the situations in which their scores were identical. We found no significant difference between the mean blame ratings for the first two athletes and the third athlete in either the certain cases ($M = 0.02$, $SD = 0.44$) or the uncertain cases ($M = -0.07$, $SD = 0.7$). Furthermore, there was also no significant difference for the credit ratings between the certain ($M = 0.28$, $SD = 1.09$) and the uncertain ($M = 0.17$, $SD = 0.68$) cases.

Discussion

As in [Experiment 1](#), the blame and credit ratings varied depending on how the athletes performed. However, in contrast to the results of [Experiment 1](#), the Certainty of Outcome factor now had only a small influence on participants' responsibility attributions for the third athlete in the nonidentical cases, and no influence in the identical cases. As mentioned above, the small effect of the Certainty factor was likely a result of the differences in relative performance in the nonidentical cases. Most of the participants were not influenced by the experienced order of events and took into account the fact that the third athlete did not know her teammates' scores. Hence, the level of the third athlete's performance was not discounted in situations in which the outcome was already determined. In addition, whether the events actually happened simultaneously or sequentially did not affect attributions.

General discussion

The results of two experiments showed that people exhibit an attenuation effect in their responsibility attributions for team members contributing sequentially to a team challenge. As predicted by the CCM, the level of a team member's performance is discounted for situations in which the outcome is already determined. Importantly, this effect is only present in situations in which the later team member knows the previous members' scores, and it is not affected by the objective order of events. The effect does not overgeneralize to situations in which the order in which participants learn about the scores does not map onto the order in which they are generated: The attenuation effect is moderated by the inferred epistemic states of the athletes.

In both experiments, attributions of responsibility were also strongly affected by the level of performance: Athletes generally received more credit and less blame for a high score than for a low score. Importantly, the effect of performance was present even in situations in which the team result was already certain and known to the athletes. We take this finding as evidence that participants' responsibility attributions are not solely determined by whether an individual's contribution made a difference in the actual situation.

The CCM can be extended in a natural way to capture the patterns of responsibility attribution from our study. Rather than determining an individual's causal responsibility solely in terms of the difference her contribution made to the team outcome in the actual situation, it also matters whether her contribution was likely to have made a difference in other possible situations—namely, situations in which the team result was not yet certain prior to her contribution.

We do not think that participants' sensitivity to differences in performance in situations in which the result was already certain is unreasonable. First, as outlined in the introduction, the level of performance of the third athlete conveys important information. If, for example, an athlete performed well despite the fact that the team had already lost for sure, we would learn that the athlete was in principle capable of good performance. This good performance could have made a difference, in the counterfactual situation in which the teammates had performed somewhat better. In contrast, if the third athlete performed poorly, we could not be sure whether this was due only to the fact that she knew that the team had already lost, and hence did not try hard, or whether she could not have performed better even if the contribution had been needed.

Second, taking an individual's performance into account also makes sense if one considers the forward-looking motivational function that attributions of responsibility can serve (cf. Weiner, 1995): Receiving blame for a negative outcome or credit for a positive outcome motivates better performance in the future. Despite the fact that how one performed might not have been critical in the present situation, a high performance might well make the crucial difference in a future situation.

Our theoretical extension of the CCM makes straightforward predictions: The degree to which a contribution is judged as being responsible for the outcome will not only increase as a function of whether it made a difference in the actual situation, but also with an increased chance that it would have made a difference in similar possible situations. Conversely, responsibility is expected to be low when a contribution does not make a difference in the actual situation and is also unlikely to make a difference in other possible situations (cf. Wells & Gavanski, 1989).

Our proposal is in line with recent work that has highlighted the close relationship between counterfactuals and attributions of causality and responsibility (Chockler & Halpern, 2004; Halpern & Pearl, 2005; Petrocelli, Percy, Sherman, & Tormala, 2011). According to Chockler and Halpern's structural model of responsibility, an agent is maximally responsible if she made a difference in the actual situation. However, responsibility does not immediately drop to zero when an agent's contribution made no difference. Rather, an agent's responsibility decreases as a function of the distance of the actual situation from a situation in which her contribution would have made a difference. One of the important advantages of this approach is that it can deal with cases of causal overdetermination, in which there are multiple sufficient causes. While each individual cause actually makes no difference to the outcome in such a situation, there is a possible situation for each cause in which it would have made a difference—namely, the situation in which the other causes were absent. The further away such a situation is from the actual situation (i.e., the more changes would need to be made), the less responsible is the cause under

consideration. For further discussion of the empirical applicability of Chockler and Halpern's model, see Lagnado, Gerstenberg, and Zultan (2012), and Zultan, Gerstenberg, and Lagnado (2012).

While the results of this experiment support the proposed extension of Spellman's (1997) model, alternative explanations cannot be ruled out on the basis of the data from the reported experiments alone. For example, one might argue that participants' responsibility attributions are mostly determined by the athletes' scores and that the attributions to the third athlete in the first experiment were subject to an anchoring-and-adjustment process. While such a performance-adjustment account could explain the results for the reported experiments, in which individual performance was measured by a discrete variable, such an account would have difficulty capturing situations in which performance was measured by binary variables. Indeed, Zultan, Gerstenberg, and Lagnado (2012) have shown that in situations in which each team member's performance is binary (i.e., they can either succeed or fail in their individual task), responsibility attributions for the team outcome are still sensitive to how close an individual's contribution was to making a difference. Another restriction in the present set of experiments was that the players all performed identical tasks. It would be natural to extend the model to situations in which the individual tasks differ in their difficulty, or their criticality for the team outcome.

In sum, our experiments highlight the fact that a comprehensive model of responsibility attribution in group contexts will need to take into account the individuals' mental states (Gerstenberg, Lagnado, & Kareev, 2010), the extent to which each contribution made a difference to the outcome in the actual situation, and whether the contribution could have made a difference if things had turned out somewhat differently (Chockler & Halpern, 2004; Gerstenberg & Lagnado, 2010).

Author note T.G. is the beneficiary of a doctoral grant from the AXA Research Fund. D.A.L. was supported by ESRC Grant RES-062-33-0004. We gratefully acknowledge the contributions of Maarten Speekenbrink and Catherine Cheung. We also thank three anonymous reviewers for comments on a previous version of this article. Part of this work was presented at the 33rd Annual Meeting of the Cognitive Science Society and is published in the conference proceedings.

References

- Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research*, 22, 93–115.
- Gerstenberg, T., & Lagnado, D. (2010). Spreading the blame: The allocation of responsibility amongst multiple agents. *Cognition*, 115, 166–171.
- Gerstenberg, T., Lagnado, D. A., & Kareev, Y. (2010). The dice are cast: The role of intended versus actual contributions in responsibility attribution. In S. Ohlsson & R. Catrambone (Eds.), *Cognition in flux: Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 1697–1702). Austin, TX: Cognitive Science Society.
- Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science*, 56, 843–887.
- Hilton, D. J., McClure, J., & Sutton, R. M. (2010). Selecting explanations from causal chains: Do statistical principles explain preferences for voluntary causes. *European Journal of Social Psychology*, 40, 383–400.
- Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, 108, 754–770.
- Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2012). *Causal responsibility and counterfactuals*. Manuscript submitted for publication.
- Mandel, D. R. (2003). Judgment dissociation theory: An analysis of differences in causal, counterfactual and covariational reasoning. *Journal of Experimental Psychology: General*, 132, 419–434.
- Miller, D. T., & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of Personality and Social Psychology*, 59, 1111–1118.
- Petrocelli, J. V., Percy, E. J., Sherman, S. J., & Tormala, Z. L. (2011). Counterfactual potency. *Journal of Personality and Social Psychology*, 100, 30–46.
- Robbennolt, J. K. (2000). Outcome severity and judgments of "responsibility": A meta-analytic review. *Journal of Applied Social Psychology*, 30, 2575–2609.
- Spellman, B. (1997). Crediting causality. *Journal of Experimental Psychology: General*, 126, 323–348.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.
- Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, 56, 161.
- Wells, G. L., Taylor, B. R., & Turtle, J. W. (1987). The undoing of scenarios. *Journal of Personality and Social Psychology*, 53, 421–430.
- Zultan, R., Gerstenberg, T., & Lagnado, D. A. (2012). *Finding fault: Counterfactuals and causality in group attributions*. Manuscript in revision.