Blame the Skilled

Tobias Gerstenberg (t.gerstenberg@ucl.ac.uk)[1], Anastasia Ejova (anastasia.ejova@adelaide.edu.au)[2], David A. Lagnado (d.lagnado@ucl.ac.uk) [1]

[1]Department of Cognitive, Perceptual, and Brain Sciences
University College London, United Kingdom

[2] School of Psychology, University of Adelaide, Australia

Abstract

This study investigates the influence of players' performance and level of skill on responsibility attributions in groups. Participants act as external judges and evaluate the performance of teams of differently skilled players who compete in a darts game. The results show that both performance and skill influence responsibility attributions. Poor performance elicits high blame and low credit ratings and vice versa for good performance. Responsibility attributions to one player did not vary as a function of the other player's performance. The influence of skill on responsibility attributions was asymmetric. While skilled players were blamed more for losses than unskilled players, credit ratings did not vary significantly as a function of skill. This result is in line with people's strong tendency to spontaneously consider upwards counterfactual alternatives for losses over downwards counterfactuals for wins.

Keywords: responsibility attribution; counterfactual thinking; control; skill.

Introduction

Consider you witness the following situation on a night out in a pub. Three friends are playing darts and, to spice things up, one of them offers the other two the following deal: "Both of you throw at the same time. If one of you manages to hit the dart in the center region, the next round of drinks will be on me. However, if none of you hits the center, you'll have to pay for my next pint."

You have seen from their previous play that one of the players is very skilled. In fact, she managed to hit the center region most of the time. The other player's performance, in contrast, was quite poor. He hardly ever managed to get the dart in the center. How would you spread the blame if neither of them managed to hit the center? Whom would you credit more if both of them hit the center? This paper investigates how people attribute responsibility between multiple agents based on their underlying skill level and actual performance.

Skill, Expectation and Control

The problem of how credit for a positive outcome or blame for a negative outcome should be distributed across the members of a group is encountered in many contexts – from law, business and medicine, to heated dinner table debates about team sports. Skill and performance are important variables that potentially differentiate the individual agents contributing to a joint effort and are hence likely to influence credit and blame attributions. How skilled we think a person is has a direct influence on what performance we expect from her. Furthermore, skill is closely connected to the notion of control. If a person is skilled it implies she is able to do something well in a reliable fashion. However, as the well-known phenomenon of choking in sports demonstrates, a player might fail to deliver because he struggles with the external pressure imposed by high expectations. Hence, high skill does not necessarily imply good performance. Similarly, a low skilled person can sometimes surprise with a very good performance. How do considerations about skill and performance influence people's attributions of blame or credit and what cognitive processes are likely to guide responsibility attributions in these contexts?

Achievement Motivation: Ability and Effort

A rich literature in attribution research has been concerned with analyzing the causal factors that are perceived to influence an agent's success or failure in achievement related contexts (see e.g. Weiner, 1995). In one of the very first studies, Weiner and Kukla (1970, Experiment 1) presented scenarios in which they systematically varied the ability and effort of hypothetical students paired with different performance outcomes. For example, a student could be described as having low ability, expended high effort and achieved an excellent grade in their exam. Based on this information, participants were asked to assign reward or punishment to the students. Whether students received punishment or reward was directly related to the outcome of their exam whereby participants showed a tendency to reward more than punish. Additionally, participants' responses were significantly influenced by both ability and effort. Students who expended high effort were rewarded more and punished less than students who expended low effort. Furthermore, students with high ability received more punishment and less reward compared to students with low ability. Interestingly, whereas both able and non-able students received the same reward for the best possible outcome (an excellent exam), able students received more punishment than non-able students for the worst possible outcome (a clear failure in the exam). Overall, however, reward and punishment were more strongly influenced by differences in the expended effort than ability of the students.

In order to explain this difference, *controllability* has been identified as an important factor that dissociates effort from ability (Weiner, 1995). In Alicke's (2000) model of personal control, a useful distinction is drawn between *behavior control* and *outcome control*. Whereas how much effort we expend is a behavior we have control over, we cannot behaviorally control our ability at a given moment. How much causal control a person has over an outcome, however, depends to a large extent on the person's ability

(as well as her effort). As mentioned above, a person's outcome control increases with her skill. In this paper we will primarily be interested in the effects that perceived *outcome control* has on a person's responsibility. The degree to which a person possesses outcome control not only depends on her capacities but also on counterfactual considerations about whether the outcome would have been different had she acted differently (Wells & Gavanski, 1989).

Counterfactual Thinking and Causal Inference

Counterfactual thoughts are thoughts about alternative events in the past and the hypothetical future outcomes they would have resulted in. For example, if a student failed their exam she might think about what she could have done differently (e.g. study instead of going to the beach) so that she would have passed the exam. Counterfactual thoughts can be distinguished by their directionality of contrast. *Upward* counterfactuals are comparisons of the actual world with a somewhat better world and *downward* counterfactuals involve the supposition of a worse world. Several studies have shown that people are more likely to spontaneously engage in upward counterfactual thinking (e.g. Sanna & Turley, 1996). Downward counterfactuals, in contrast, are endorsed comparatively rarely. Accordingly, an outcome's valence – or, more specifically, the affective state motivated by the valence – is one of the main determinants for the activation of the counterfactual thinking process (Roese, 1997). Apart from an outcome's valence, the degree to which the outcome was to be expected has been identified as a promoter for spontaneous causal (Kanazawa, 1992) and counterfactual thoughts (Kahneman & Miller, 1986; Sanna & Turley, 1996). The less an outcome was expected the more likely were people to engage in causal or counterfactual thinking.

Counterfactuals and Responsibility

Several researchers have argued for the close relationship between counterfactuals, causation and responsibility attribution (Hilton & Slugoski, 1986; Shaver, 1985). In law, the *but for* rule is a standard criterion for identifying a person's action as the cause-in-fact, which is a precondition for the person to be held responsible for the negative event. Accordingly, a person can only be held liable if the negative event would not have come about *but for* his action. Wells and Gavanski (1989) have demonstrated that counterfactual alternatives indeed influence people's ratings of causality. A person was rated more causal for a negative outcome, if the outcome could have been prevented had he acted differently, compared to a situation in which the outcome would have occurred even if he had acted differently.

In situations in which there are multiple people involved, a person's control over the outcome is not exclusively determined by their own skill but also by the other people's abilities as well as the way in which the individual contributions are combined to determine the outcome. Gerstenberg and Lagnado (2010) have shown that the same performance can be evaluated differently depending on the group task and the performance of the other players. Their paper provided the first empirical test of a structural model of responsibility attribution developed by Chockler and Halpern (2004). At the core of this model is a relaxed notion of counterfactual dependence, according to which an event can still be identified as a cause even if changing it would not have made a difference to the outcome in the actual situation. In their model, an individual agent's responsibility for a group's outcome equals $1/(N+1)$, whereby N denotes the minimal number of changes from the actual situation that would have been necessary to generate a situation in which that agent's contribution would have made a difference to the outcome. If no change is needed, the agent receives a responsibility of 1. The more changes would have been necessary to make a person's contribution critical, the more her responsibility decreases.

Consider, for example, our initial darts scenario. In order for the two friends to win the bet, at least one of them needs to hit the center region. In a situation in which both players hit the center, their win is overdetermined. That is, the outcome does not depend on either of the players' individual action and hence, a simple *but for* counterfactual analysis would not identify either of them as a cause for the positive outcome. Each player's contribution would only have made a difference to the outcome, if the other player had not hit the center. Expressed in terms of the structural model of responsibility attribution each person required one change from the actual situation to be critical and should hence receive a responsibility of 1/2. Thus, the model predicts that a player's credit should be reduced if the other player hit the center as well.

Experiment

In order to assess how people attribute blame and credit in a group setting as a function of the players' underlying skill and actual performance, we used the context of a game show environment in which players participated in a team challenge whose outcome affected their individual payoff. The game was similar to an ordinary darts game (see Figure 1). It consisted of two phases: First, there was a practice phase in which each player was given 20 practice shots. Second, there was the crucial team challenge in which two players were put together randomly to form a team. The team won their challenge if at least one of the two players managed to hit the dart in the center region. Participants were told that the players differed in terms of how skilled they were in the task. The practice shot patterns were used to manipulate the players' skill levels (see Materials).

The participant's task was to indicate to what extent each player was responsible for the team's result. Participants attributed blame to each player if the team lost and credit if it won. They were informed that their ratings would affect the player's payoff. The more blame a player received for the team's loss, the more his payoff was reduced. The more credit a player received for her team's win, the more her payoff was increased.

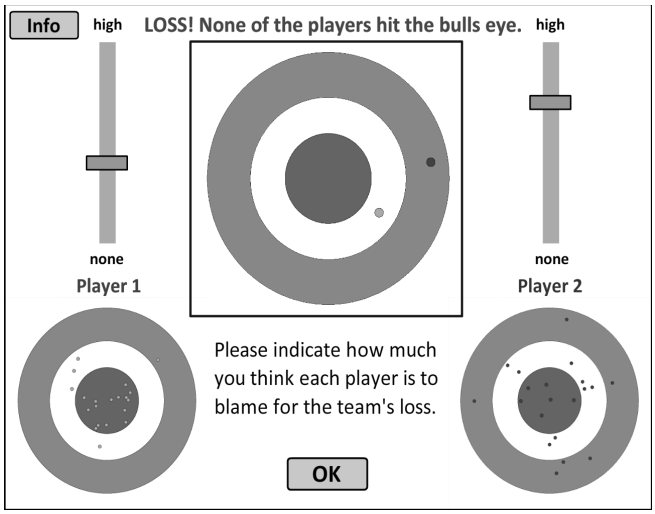


Figure 1. Screenshot of the game. Each player's performance in the practice is shown in the bottom corners. The performance in the crucial team challenge is shown in the center.

We hypothesized that both a player's perceived outcome control as well as the actual performance would influence participants' responsibility attributions. The control factor, as manipulated through the player's skill level, directly influences the availability of counterfactual alternatives (Girotto, Legrenzi & Rizzo 1991). If a skilled player missed the center region, the alternative event in which she did hit the center is highly available. Likewise, if an unskilled player hits the center, the alternative event in which he misses is readily available.

Given the prevalence of spontaneous upwards rather than downwards counterfactuals, we expected the influence of counterfactual alternatives to be stronger in the case of outcomes with negative valence. Accordingly, the skilled player would be blamed more for a loss than the unskilled player. Since downward counterfactual thinking has rarely been shown to occur spontaneously, we expected only a small influence of the skill factor for outcomes with positive valence.

Furthermore, we expected that participants' blame and credit ratings would vary as a function of actual performance. Despite the fact that the rule of the game employs a clear cut-off point in that it only matters whether a player hits the center region or not, we expected that blame ratings for losses would increase with an increased distance of a shot from the center. For wins we expected that players would receive most credit if they hit the center. Furthermore, we expected that players would receive only minimal credit if they did not hit the center and that credit ratings would be higher the closer they were to the center.

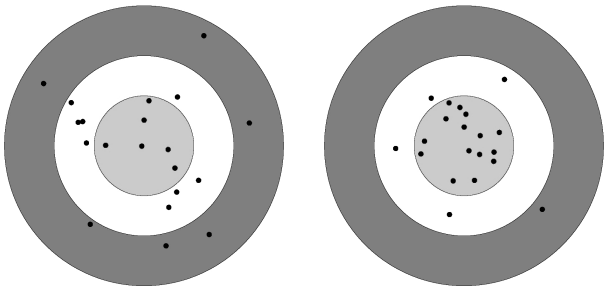
Finally, taking the considerations of the structural model of responsibility attribution into account, we expected that a player's responsibility rating would vary as a function of the other player's performance. More specifically, we expected a player's credit rating to be reduced for cases in which the outcome was overdetermined. Hence, a player should receive less credit if the other player also hit the center as compared to situations in which the other player missed.

Method

Participants 52 participants (31 female) were recruited through the UCL subject pool and took part to receive course credit points or for the chance of winning Amazon vouchers worth £60 in total. The mean age was 23.9 (*SD* = 6.3).

Design The experiment employed a 3 (*skill levels*: both players unskilled, both player skilled, one player skilled and one player unskilled) x 9 (*performance patterns*: full permutation of 3^2 possible shots (center, medium, outside region) for pairs of players) within-subjects design (see Figure 3). Given that the team wins if at least one of the players hits the center region, this design resulted in 15 cases in which a team won and 12 cases in which they lost.

Materials Figure 2 shows an example for the practice shot pattern of the unskilled player and the skilled player. A prototype was generated for each skill level by sampling 20 data points from two centered independent Gaussian normal distributions for the x-axis and y-axis. The skill was manipulated by varying the variance of the distribution. For the unskilled player pattern, 6 shots hit the center, 8 shots the middle, and 6 shots the outside region. For the skilled player, 15 shots hit the center, 4 shots the middle and 1 shot the outside region. Hence, based on the practice pattern, the unskilled player had a 30% chance and the skilled player had a 75% chance of hitting the center. From the prototypical skill patterns, we generated 27 patterns each by independently rotating the individual shots around the center. This procedure ensured that the practice patterns of different players with the same skill level were matched with respect to the most important characteristics. The summed distance of the shots to the center as well as the number of shots in the different regions was held constant. Nevertheless, the practice patterns still looked different between the players. The patterns of shots for the crucial team challenge were created in a similar fashion. One prototypical center, middle and outside ring shot was created and randomly rotated for each pattern in the experiment. Hence, the actual distance of a center, middle or



unskilled player **skilled player**

Figure 2. Prototypical practice shot patterns for the unskilled and skilled player.

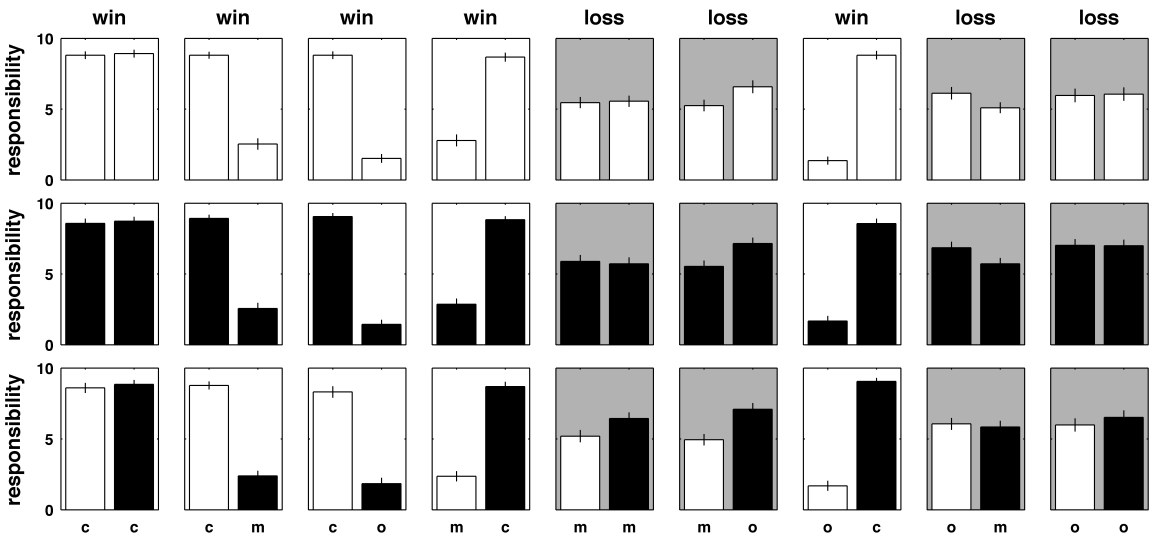


Figure 3. Mean credit and blame attributions (±*SE*) for the 27 patterns used in the experiment. White bars = unskilled player, black bars = skilled player; c = shot in the center region, m = middle region, o = outside region.

outside shot was identical in each of the combinations.

Procedure The study was carried out online.[1] At the beginning of the experiment, participants were instructed that they would take the role of a judge in a game show and that their task will be to evaluate players' performances. The nature of the practice trials and the rules for the team challenge were described as explained above. Participants were told that the players in the game show differed with respect to how skilled they were. As a manipulation check, we showed them 3 patterns of practice shots after initial instructions and asked them to indicate how skillful each of the players was in the task. Participants used a slider for each skill pattern ranging from -10 (very unskilled) to +10 (very skilled). The mean ratings for the unskilled player were M = -3.1 (SD = 3.2), for the medium skilled player M = 1.4 (SD = *2.5*) and for the skilled player M = 6.1 (SD = 3). In the main part of the experiment, only the patterns of the unskilled and skilled player were used.

After the skill manipulation check, participants did one practice trial in which the different components of the screen were explained. By clicking on an 'Info' button which remained on the screen throughout the experiment, participants could always remind themselves of the most important aspects of the task. After the practice trial, participants answered a series of 4 forced choice comprehension check questions. On average, participants answered 75% of the questions correctly. After having given an answer, the correct solution was displayed. Participants then proceeded to the main stage of the experiment, in which they evaluated the performance of 27 teams of different players. They always saw each player's performance in the practice trials first and then the result in the team challenge was revealed. If one of the two players hit the center region, the team won the challenge, otherwise they lost. Participants were informed about the result of the challenge at the top of the screen. To identify the different players, their shots were colored differently. If the team won the challenge, participants attributed credit to each player. If the team lost the challenge, participants attributed blame. The sliders ranged from 0 ('none) to 10 ('high').

At the end of the experiment, participants saw the practice patterns and shots in the team challenge for 4 individual players sequentially. Two players were skilled and two players were unskilled. For each of the skill levels, one of the players hit the center and one of the players hit the outside ring. For each of the 4 patterns, participants were asked to indicate how much the following factors influenced the player's result on the final test shot. The factors were: 'The player's skill level', 'The player's effort', 'The pressure of the situation', 'Chance' and 'The intention to perform this shot'. Participants made their ratings on separate sliders ranging from 0 ('not at all') to 10 ('very much'). This final stage was used to gain insight into how participants might explain the different results based on the factors provided. Finally, participants were asked to provide their age and gender.

Results

For all statistical tests, we have adopted a significance criterion of $p < .05$. Blame ratings for losses and credit ratings for wins were analyzed separately.

Figure 3 shows the mean credit and blame attributions for the 27 different patterns used in the main stage of the experiment. The first row shows situations in which both players were unskilled, the second row in which they were both skilled and the third row shows the results for the mixed challenges. We analyze, in turn, the effects of actual performance and underlying level of skill on responsibility attributions.

[1] A demo of the experiment can be accessed here: http://www.ucl.ac.uk/lagnado-lab/research.html

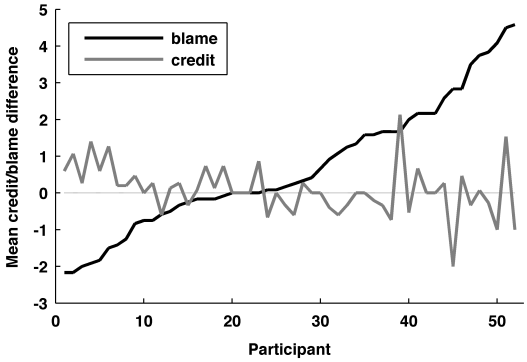


Figure 4. Individual differences in the effect of skill on blame/credit attributions. Positive values = skilled player is favored, negative values = unskilled player is favored.

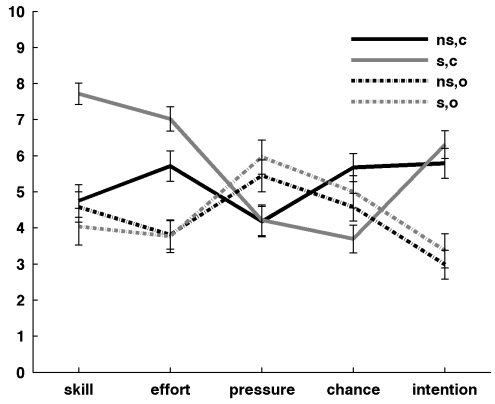


Figure 5. Mean ratings (±*SE*) indicating how much different factors were seen as having contributed to a shot. ns = non-skilled, s = skilled, c = center, o = outside.

The Influence of Performance First, we wanted to see whether blame and credit ratings varied as a function of performance. Indeed, credit ratings were significantly influenced by performance, $F(1,51) = 428.18$, $\eta = .894$. Players received most credit if they hit the center region ($M = 8.77$, $SD = 1.39$). Furthermore, credit ratings for player that did not hit the center (M = 2.09, SD = 1.81) were significantly greater than 0, $t(51) = 8.32$. Players received significantly more credit if they hit the middle ring ($M = 2.58$, $SD = 2.04$) compared to the outside ring ($M = 1.59$, $SD = 1.75$), $t(51) = 6.07$. Similarly, blame ratings were influenced by the performance of the player as well. A player received more blame if she hit the outside ($M = 6.53$, $SD = 2.23$) compared to the middle ring ($M = 5.55$, $SD = 2.16$), $t(51) = -4.94$.

To test how a player was evaluated depending on the performance of the other player, we compared how much credit a player received for a shot in the center region if the other player also hit the center or not. A player's credit for a center shot if the other player also hit the center ($M = 8.75$, $SD = 1.67$) was not significantly different from situations in which the other player did not hit the center ($M = 8.77$, $SD = 1.34$).

The Influence of Skill Second, we wanted to see whether the blame and credit ratings differed as a function of the player's skill levels. Overall, the skilled players received more blame for the team's loss ($M = 6.4$, $SD = 2.23$) than the unskilled players ($M = 5.69$, $SD = 2.29$), $t(51) = -2.87$. However, there was no significant difference between the credit ratings for skilled players ($M = 6.13$, $SD = 1.07$) and unskilled players ($M = 6.05$, $SD = 0.98$), $t(51) = -0.88$.

To look more closely at the effect that the skill level had on people's attributions, we compared the situations in which both player's performance was identical but their skill differed. Table 1 shows that the proportions of participants that either gave equal ratings to both players in these cases or favored one player over the other differed significantly, $\chi^2(4, N = 52) = 16.83$. The majority of participants attributed credit equally when both players hit the center. However, in situations in which the team lost and both players either hit the middle or the outside ring, a majority of participants assigned more blame to the skilled compared to the unskilled player.

Figure 4 shows the effect of the skill manipulation for each individual participant: Shown are the mean differences in blame and credit attributions between skilled and unskilled players. Positive differences mean that a participant attributed more blame/credit to the higher skilled player. For losses, 29 participants attributed more blame to the skilled player, 19 participants less and 4 participants gave equal blame. For wins, 17 participants attributed more credit to the skilled player, 24 participants less and 11 participants gave equal credit. Participants' blame and credit attributions were negatively correlated as a function of skill, $r = -.34$. Hence, the more blame a participant attributed to a *skilled* player compared to an *unskilled* player, the more she credited the *unskilled* player compared to the *skilled* player.

Figure 5 shows to what extent participants perceived different factors to be important in explaining the players' results for the 4 test cases at the end of the experiment. We will only discuss the results descriptively. Participants considered the 'Skill' factor to be most important for explaining the shot in the center by the skilled player. 'Effort', 'Pressure' and 'Intention' varied as a function of performance. For good performances, participants assumed that the player put in high effort, resisted the pressure of the situation and intended to bring about the outcome. The reverse pattern was found for bad performances. The 'Chance' factor varied as a function of expectation. The mean rating for the skilled person hitting the center was lowest and the rating for the unskilled person hitting the center highest.

Table 1. Proportions of participants that either gave identical ratings in mixed-skill challenges with identical performance, favored the unskilled or skilled player.

	identical	**unskilled**	**skilled**
center	34	9	9
middle	17	9	26
outside	18	12	22

Discussion

The results revealed that the quality of performance influenced both blame ratings for losses and credit ratings for wins. The worse a person performed the more blame he received for the loss and the less credit for a win. Players received marginal credit for the team's win even if they did not hit the center region. How a player's performance was evaluated did not vary as a function of the teammate's performance. The influence of skill on responsibility attributions was asymmetric. Skilled players received more blame than unskilled players for losses but credit attributions for wins did not differ significantly as a function of skill.

General Discussion

In a novel paradigm in which we systematically varied the skill and performance of agents in a group task, we found that both factors significantly influenced participants' responsibility attributions.

While the finding that responsibility attributions vary as a function of performance is quite intuitive, the fact that attributions to an individual were not affected by their teammate's performance is surprising. In a different task, Gerstenberg and Lagnado (2010) did find that participants were sensitive to the performance of the other players and the way in which individual contributions translated into the group's outcome. One important difference between the two studies concerns the reward function. While in Gerstenberg and Lagnado (2010) the team was rewarded as a whole, participants in the current study were instructed that their blame and credit ratings would affect the payoff of individual players directly. This instruction might have made participants consider the players independently and hence no reduction of credit was observed if both players performed well.

Another interesting finding concerns the asymmetric effect of the skill manipulation on participants' responsibility ratings. This partly replicates Weiner and Kukla's (1970) finding that reward did not vary as a function of ability (at least for very good outcomes) but punishment did. It is also in line with previous research in the counterfactual literature that showed that counterfactual thoughts are more likely to be spontaneously elicited for outcomes with negative as opposed to positive valence. The fact that the counterfactual alternative in which the skilled player, who exerts more control over the outcome, hit the center region is more easily available explains the increased blame ratings in these situations. If violations of expectation were the main driving force of attributions independent of the valence of the outcome, one would have also expected an increased credit rating for the unskilled player. However, our asymmetric results rule out this explanation.

It is likely that the influence of the skill manipulation would have been even stronger if we had chosen a sample that was representative of the player's skill levels for the patterns of shots in the team challenges. The fact that skill level in the practice and level of performance in the team challenges were independent due to our balanced experimental design, might have led some participants to disregard the skill manipulation.

One of the features of our paradigm is that the effect of different combination functions on people's responsibility attributions can be investigated. In our setup, only one of the players needed to perform well in order for the team to win. However, a situation in which both players' good performance is needed is more likely to make participants view the players as a team and hence stronger effects of one player's skill and performance on the other player's evaluation are to be expected.

Acknowledgements

TG is the beneficiary of a doctoral grant from the AXA research fund. AE was supported by an APA scholarship, DL was supported by ESRC grant (RES-062-33-0004).

References

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, 126 (4), 556-574.

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. A. (2010). Spreading the blame: The attribution of responsibility amongst multiple agents. *Cognition*, 115, 166-171.

Girotto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica*, 78, 111-133.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, 93 (1), 75-88.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93 (2), 136-153.

Kanazawa, S. (1992). Outcome or expectancy? Antecedent of spontaneous causal attribution. *Personality and Social Psychology Bulletin*, 18 (6), 659-668.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin*, 121 (1), 133-148.

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.

Shaver, K. G. (1985). *The Attribution of Blame: Causality, Responsibility and Blameworthiness*. New York: Springer-Verlag.

Weiner, B. (1995). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92 (4), 548-573.

Weiner, B., and Kukla, A. (1970). An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology*, 15, 1–20.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology*, 56 (2), 161-169.