Social Cognitive and Affective Neuroscience, 2017, 138–145

doi: 10.1093/scan/nsw160
Advance Access Publication Date: 1 December 2016
Original article

Beyond self-serving bias: diffusion of responsibility reduces sense of agency and outcome monitoring

Frederike Beyer,[1,*] Nura Sidarus,[1,2,*] Sofia Bonicalzi,[1,3] and Patrick Haggard[1,3]

[1]Institute of Cognitive Neuroscience, University College London, London, UK, [2]Institut Jean Nicod, Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL Research University, 29 rue d'Ulm, 75005 Paris, France, and [3]School of Advanced Study, University of London, London, UK

Correspondence should be addressed to Frederike Beyer, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AZ, UK. E-mail: f.beyer@ucl.ac.uk

*Equally contributing authors.

Abstract

Diffusion of responsibility across agents has been proposed to underlie decreased helping and increased aggression in group behaviour. However, few studies have directly investigated effects of the presence of other people on how we experience the consequences of our actions. This EEG study investigated whether diffusion of responsibility simply reflects a *post-hoc* self-serving bias, or rather has direct effects on how we process the outcomes of our actions, and our experience of agency over them. Participants made voluntary actions whose outcomes were more or less negative. Presence of another potential agent reduced participants' sense of agency over those outcomes, even though it was always obvious who caused each outcome. Further, presence of another agent reduced the amplitude of feedback-related negativity evoked by outcome stimuli, suggesting reduced outcome monitoring. The presence of other agents may lead to diffusion of responsibility by weakening the neural linkage between one's actions and their outcomes.

Key words: diffusion of responsibility; sense of agency; social interaction; FRN; outcome monitoring

Introduction

Social psychology has long recognised that the presence of other people substantially influences behaviour. Perhaps the most troubling example is the so-called 'bystander effect': the presence of others reduces the likelihood that people will help in an emergency situation or interfere with social norm violations (Darley and Latane, 1968; Chekroun and Brauer, 2002). A much more common, and socially problematic effect is 'social loafing'. When a group of people has to work towards a collective goal, each individual on average puts in less effort than they would when working alone (Karau and Williams, 1993). Experimental studies also show that groups tend to make riskier choices than individuals (Wallach *et al.*, 1964; Bradley, 1995), and behave more aggressively (Bandura *et al.*, 1975; Meier and Hinsz, 2004). All these situations have in common that individual behaviour is altered in social contexts. The presence of other people makes agents feel less responsible for the outcome of group decisions, especially those with negative consequences (Mynatt and Sherman, 1975; Forsyth *et al.*, 2002).

These findings have led to the concept of 'diffusion of responsibility': the idea that the presence of others changes the behaviour of the individual by making them feel less responsible for the consequences of their actions (Bandura, 1991). The diffusion of responsibility concept has great social, political and moral importance, because it might constitute a form of moral disengagement purported to explain inhumane actions (Bandura, 1999). However, it remains unclear whether the mere presence of others actually changes the experience of action

Received: 29 June 2016; **Revised:** 15 September 2016; **Accepted:** 17 October 2016

and responsibility, or merely triggers a *post-hoc* bias in reports of responsibility, to preserve self-esteem. In order to play a causal role in group behaviour, diffusion of responsibility would need to have 'online' influences on how people experience a given situation, and not merely constitute a *post-hoc* narrative that individuals can use to explain outcomes after the fact. Few previous studies have focussed on potential 'online' mechanisms by which the presence of other agents could influence the experience of action.

Sense of agency refers to the feeling that one can control external events through one's own actions. Sense of agency plays a crucial role in social interactions (Frith, 2014), and is therefore tightly linked to the experience and allocation of responsibility. Besides explicit self-reports of sense of agency, a more objective, and implicit, measure of action-outcome processing can be obtained using event-related potentials (ERPs). The feedback-related negativity (FRN) is an ERP-component associated with monitoring the consequences of action (see San Martín, 2012 for a review). Importantly, this component is sensitive to the perceived controllability of action outcomes: when participants believe that an outcome is uncontrollable, the FRN to negative outcomes is greatly reduced (Yeung *et al.*, 2005; Li *et al.*, 2011). The FRN is also sensitive to the motivational significance of outcomes (Gehring and Willoughby, 2002; Holroyd and Yeung, 2012), potentially explaining the inverse relation between controllability and FRN amplitude. Uncontrollable outcomes are less important to the agent, as they provide little information on how to improve behaviour.

The presence of others may reduce sense of agency through increased authorship ambiguity and an objective decrease in control. For example, a joint grade for a group project provides little information about the quality of individual contributions. Accordingly, Li *et al.* (2010) showed that in a dice-tossing task, FRN amplitude was reduced when, instead of tossing all three dice, participants tossed only one, while the other dice were tossed by other players. Therefore, the presence of other players seemingly reduced participants' control over the outcome by two-thirds. However, diffusion of responsibility occurs even when control is unaffected by the presence of others. In the classic 'bystander effect' (Darley and Latane, 1968), the fact that several people witness an emergency does not undermine the capacity of one person to act and alter events. Thus, to explain why the presence of others changes people's behaviour, diffusion of responsibility would have to influence an individual's experience of the situation, beyond objective effects on action-outcome contingencies. Surprisingly, this possibility has been largely neglected in the literature.

We propose that this reduction in sense of agency may be mediated by the complexity of social decision-making compared with individual decision-making. Difficulty, or dysfluency, in decision-making has been shown to reduce sense of agency for the outcome of the decision (for a review, see Chambon *et al.*, 2014). In social situations, one needs to consider the potential actions of others. This makes action selection more difficult. This complexity during 'action selection' might then affect the processing of action outcomes, even if the outcome monitoring itself is no more complex or demanding in social compared with non-social situations.

We investigated whether diffusion of responsibility might arise because the individual sense of agency over actions and outcomes is automatically reduced in the presence of alternative agents. Importantly, this social dilution of agency should not simply reflect 'ambiguity' about who is responsible for the outcome, nor changes in action-outcome contingencies. Rather, it should represent a reduction in the impact or significance of action outcomes in social *vs* non-social settings. To this end, we designed an experiment with two agency conditions that differed only in terms of social context. This required: (i) action consequences to be controllable, and (ii) attribution of outcomes to the participant's own actions to be unambiguous in both the social and non-social context. Previous studies involved objective decreases in control over outcomes, by eliminating response choices (Yeung *et al.*, 2005) or by having others act in addition to the participants (Li *et al.*, 2010). In contrast, our goal was to ensure that participants had 'objectively' the same amount of control in social and non-social contexts, thus we designed a task in which action-outcome contingencies were stable across the experiment, and participants were instructed accordingly.

We used a task in which a marble rolls down a bar, and an action is required to stop it from crashing (Schel *et al.*, 2014). In our version of this task, participants either played alone or allegedly together with another player. If the participant acted, the marble stopped immediately, so they could unambiguously attribute the outcome to their own action. As the diffusion of responsibility concept is mostly used to explain behaviour in situations where acting is somehow costly or effortful, or results in negative consequences, we designed the task to exclusively produce negative outcomes. Stopping the marble incurred some cost for the participant, but this cost was avoided if the other player stopped the marble. ERPs were recorded in response to the outcome: visual feedback of the points lost on each trial. Importantly, outcome presentation was separated in time from the immediate action feedback of the marble stopping. Therefore, participants already knew whether they would lose some points due to their own action before the outcome was presented. This ensured that the complexity of the outcome processing phase was not affected by the presence or absence of the other player. Participants then rated how much control they felt over the outcome using a visual analogue scale. Finally, our analyses focused on trials in which participants successfully stopped the marble, so that the 'only' difference between social conditions was whether a potential alternative agent was present or not.

We predicted that the presence of an alternative agent would not only influence behaviour (as participants might rely on the other player to act), but also reduce sense of agency. We further reasoned that reduced agency might either reflect a reduction in the immediate subjective experience of control, or merely a *post-hoc* justification due to a self-serving bias to blame others for one's misfortunes. The FRN component to the outcome should be reduced in the former case, but not in the latter.

Materials and methods

Participants

Previous studies investigating ERPs in relation to control over outcomes or sense of agency employed sample sizes of 16–20 participants (Li *et al.*, 2010, 2011; Kühn *et al.*, 2011; Timm *et al.*, 2014). Given that the manipulation of social context in the absence of objective condition differences has not been investigated previously, we aimed for a minimum sample size of 25. To allow for dropouts, we tested 32 healthy student volunteers (16 male, 16 female; age 18–32). The data of one participant were lost due to technical failure. The data of three other participants were excluded from data analysis because they spontaneously expressed suspicion about the co-player's participation in the task in the post-experimental questionnaire. For one participant, strong noise in the EEG signal resulted in the rejection of more than 80% of trials, and therefore this participant's data were excluded from analysis as well. Thus, data of 27 participants (12 male, 15 female) was included in the analyses.

Apparatus and materials

Participants were tested in pairs. Stimuli were presented on two identical computer screens for the two participants. Participants gave responses using standard computer mice. After the task, participants filled out a post-experimental questionnaire probing for suspicion concerning the participation of the co-player in the task, as well as the Locus of Control Scale by Rotter (1966), and the subscales 'Diffuse Responsibility' and 'Exercised Responsibility' of the Ascription of Responsibility Questionnaire (Hakstian *et al.*, 1986).

EEG was recorded from 26 channels using g.tec g.USB amplifiers with active ring electrodes and non-abrasive conductive gel. Horizontal and vertical eye movements were recorded simultaneously. EEG signals were referenced online against the left earlobe and were recorded with a 0.1 Hz Butterworth highpass filter.

Design and procedure

Participants were invited to the laboratory in mixed-gender pairs of two. They received instructions together, filled out consent forms for participation in the study and were then seated in adjoining laboratories for the testing. During the instructions, participants were assigned one avatar (Designed by Freepik.com), which would represent them during the task. They were also shown their co-player's avatar, which would be used when they played together. Both participants performed the task simultaneously, but separately. After the task was finished, participants filled out post-experimental questionnaires and personality questionnaires (see 'Materials and methods' section earlier). Participants were then fully debriefed and paid for their participation. Payment consisted of £7.50 per hour, plus any earnings from the task. To earn money from the task, participants were given monetary points at the beginning of the experiment, some of which they would lose in every trial. They were then paid according to how many points they managed to save (see task description below for details).

The marble task was designed to create a situation in which acting was costly, but withholding action was potentially more costly still. In each trial, participants had to stop a rolling marble from falling off a tilted bar, and crashing (see Figure 1). Participants were instructed that, at the beginning of each block, they would receive 1500 points worth 150 pence, and in each trial they could lose up to 100 of these points. The task consisted of 4 blocks of 30 trials each. Trials were randomly assigned to either the 'Alone' or the 'Together' condition, with 15 trials per condition and block.

In the beginning of an 'Alone' trial, participants saw their own avatar alone, indicating they would be playing by themselves, while their co-player supposedly played simultaneously on his/her computer. Next, they saw a blue marble lying on top of a tilted bar, which after 500 ms started rolling down towards the lower end of the bar. At any point, participants could press the left mouse button to stop the marble. If they did so, the marble stopped in its current position, providing immediate feedback of their successful action. If participants did not react in time, the marble rolled off the bar and crashed. The final position of the marble, whether stopped or crashed, was shown for 500ms, followed by the presentation of a fixation cross for 1500–2500 ms. In either case, participants received information about how many points they lost, i.e. the action outcome, for 2000 ms. ERPs were time-locked to outcome presentation. Afterwards, a fixation cross was presented for 500 ms and then participants saw a visual analogue scale with the question 'How much control did you feel over the outcome?' and the end points of the scale labelled 'No control' and 'Complete control'. Participants used the mouse to indicate how much control they felt they had over the number of points lost during that trial. It was emphasized during instructions that 'outcome' meant the number of points participants lost on a given trial, irrespective of whether the marble crashed.

Participants were instructed that the later they stopped the marble, the fewer points they would lose. In order to make it difficult to always stop the marble at the very end of the bar, the speed with which the marble rolled down the bar varied from trial to trial. Also, at some point along the bar, the marble would speed up, and this point varied from trial to trial. This added a risk component to the task, since if the participant waited too long, the marble might suddenly speed up and they might not be able to stop it in time to prevent a crash. There was also uncertainty about the outcome, as the exact number of points lost could not be fully predicted from the marble stopping position. In fact, the bar was divided into four different payoff sections of equal length (60–46 points at the top; 45–26 and 25–16 points in the middle; 15–1 points at the end). If the marble crashed, 70–99 points would be lost. Within each section, the number of points lost was varied randomly from trial to trial.

At the beginning of 'Together' trials, participants saw their own avatar next to the avatar of their co-player, and the marble in these trials was coloured green. Participants were instructed that, in these trials, both players would be playing together and either could use their mouse button to stop the marble. If neither player acted, the marble would crash and both players would lose the same number of points. If the co-player stopped the marble, the participant would not lose any points. If the participant stopped the marble, they would lose a number of points according to the position where they stopped it, and their co-player would not lose any points.

In fact, participants were playing alone in all trials, and the co-player's behaviour was simulated by the computer. The co-player's behaviour was programmed such that participants had to stop the marble in the majority of 'Together' trials, to ensure a sufficient number of artefact-free trials was available for ERP analyses. If participants had stopped the marble more often than their co-player, and if participants did not act sooner, the co-player could stop the marble along the lower half of the bar. In that case, the marble would stop on its own, and participants received feedback of losing zero points. To avoid ambiguity about who caused the outcome, simultaneous actions of both participant and co-player were attributed to the participant. Thus, if the participant acted within 50 ms of a simulated co-player action, this would count as participant's action, and feedback would indicate a loss according to the stop position[1].

ERP pre-processing

EEG-signals were processed using the Matlab-based open-source toolbox eeglab (Delorme and Makeig, 2004) with the ERPlab plugin (Lopez-Calderon and Luck, 2014).

The continuous EEG signal was notch-filtered and re-referenced to the averaged signal of the left and right mastoids. The signal was then cut into 3000 ms epochs time-locked to the presentation of the outcome. Independent component analysis

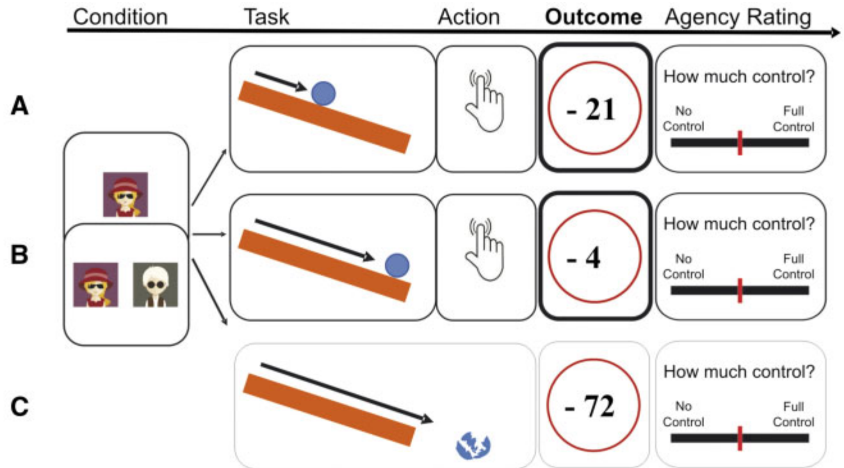

Fig. 1. Marble task. Figure shows the outline of a low-risk successful trial (**A**), a high-risk successful trial (**B**), and an unsuccessful trial (**C**). Note that C is the worst outcome, B the best, and A the intermediate. Social context was indicated at the start of a trial, by either presenting the participant's own avatar alone, or together with the other player's avatar. The marble colour served as a reminder of social context, and was either blue in the alone condition (shown here), or green in the together condition. In the together condition, besides the trials displayed here, there were trials in which the 'other' player stopped the marble, and the participant did not lose any points. ERPs were time-locked to outcome presentations of successful trials (A and B, marked in bold) in which the participant stopped the marble.

was used to remove eye movement artefacts. A 0.5 Hz highpass filter (FIR filter, cutoff frequency 0.25 Hz) and a 20 Hz lowpass filter (FIR filter, cutoff frequency 22.5 Hz) were applied. Epochs with signal artefacts were removed using an 80 μV threshold. EEG signals were then averaged into ERPs separately for the two experimental conditions, using a 100 ms pre-stimulus baseline. This resulted in an average of 39.25 (SD = 7.10) trials for the Alone condition (min = 20), and average 32.96 (SD = 9.30) trials for the Together condition (min = 16).

The FRN component was analysed as the mean amplitude between 250-330 ms, at electrode FCz, based on previous studies (Yeung *et al.*, 2005; Li *et al.*, 2011) and observation of grand ERPs and scalp topography.

Data analysis

We analysed agency ratings and ERPs only for trials where the participant acted and successfully stopped the marble. Behavioural data (stopping position, outcomes, and agency ratings) and mean FRN amplitude were analysed using hierarchical linear regression models (i.e. linear mixed-effects models). This approach is advisable with unbalanced data, and allowed us to model single trial data (Bagiella *et al.*, 2000; Baayen *et al.*, 2008; Tibon and Levy, 2015). Models included the condition as a predictor, coded as Alone = 0, Together = 1. Where relevant, Stopping Position and Outcome were also included as covariates, after standardising the values within participants. All fixed effects were also modelled as participant random effects (random intercepts and slopes). Analyses were conducted using the *lme4* package (Bates *et al.*, 2014) in R Core Team (2015). Parameter estimates (b) and their associated t-tests (t, p), calculated using the Satterthwaite approximation for degrees of freedom (Kuznetsova *et al.*, 2015), are presented to show the magnitude of the effects, with bootstrapped 95% CIs (Efron and Tibshirani, 1994).

Moreover, we analysed behavioural data (proportion of trials, agency ratings, and mean outcomes) from trials in which the marble crashed. ERP data for these trials were not analysed, however, due to low trial numbers. Finally, for together trials only, we compared the proportion of trials in which the co-player acted, relative to the marble crashing.

Results

Behaviour

The main focus of our analyses was trials in which the participant successfully stopped the marble. These trials were the same across the two social context conditions, but differed only in that participants acted while knowing that their co-player 'could have acted instead' of them in the together condition. To assess how participants' behaviour varied across social contexts, we modelled the position at which the marble was stopped. Participants stopped the marble significantly later in the together condition, relative to playing alone [$b = 3.18$, $t_{(833.30)} = 5.85$, $P < 0.001$, 95% CI = (2.12, 4.26), see Supplementary Table S1 for full results table]. This suggests that participants waited longer to act in the together condition to allow time for their co-player to act instead of them.

Outcome (number of points lost) was predicted from the social context factor, stop position covariate, and their interaction. Outcomes were related to the marble stop position [$b = 6.12$, $t_{(31.88)} = 22.54$, $P < 0.001$, 95% CI = (5.63, 6.62)], with later stops resulting in smaller losses, as expected based on the task design. The social context did not influence outcomes [$b = 0.094$, $t_{(31.98)} = 0.30$, $P = 0.77$, 95% CI = (−0.51, 0.78)], nor did the social context by stop position interaction [$b = -0.43$, $t_{(69.96)} = -1.45$, $P = 0.15$, 95% CI = (−1.00, 0.14), see Supplementary Table S2]. This shows that outcomes were similar across social contexts, for trials in which the participant successfully stopped the marble.

Finally, agency ratings were modelled using social context, stop position, and outcome, plus their interactions. Results showed a significant reduction in agency ratings when playing together, relative to playing alone [$b = -4.74$, $t_{(22.66)} = -3.57$, $P = 0.002$, 95% CI = (−7.260, −2.29); see Figure 2a]. Agency ratings were also predicted by the outcome [$b = 4.14$, $t_{(24.52)} = 4.63$, $P < 0.001$, 95% CI = (2.21, 5.61)], with smaller losses being associated with higher ratings (see Supplementary Figure S1a). Finally, agency ratings were significantly influenced by the marble stopping position [$b = 2.73$, $t_{(22.66)} = 3.03$, $P = 0.006$, 95% CI = (0.77, 4.65)], with later stops being linked to higher ratings (see Supplementary Figure S1b). There were no significant interactions (see Supplementary Table S3).

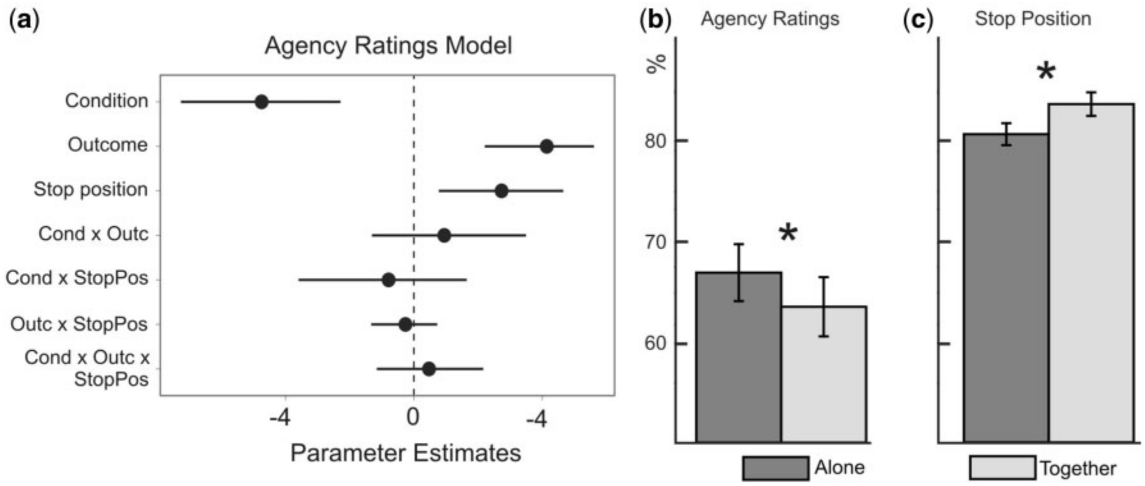

Fig. 2. Behavioural results. **(a)** Parameter estimates for the model predicting agency ratings, with 95% bootstrapped confidence intervals. *Condition* refers to the effect of social context (Alone = 0 *vs* Together = 1), such that a negative parameter estimate denotes a loss of agency in the Together condition. **(b)** Mean agency ratings for the two experimental conditions, showing a significant reduction in agency ratings in Together trials. **(c)** Mean position at which participants stopped the marble for the two experimental conditions, showing a significant delay of actions in Together trials. Error bars show standard error of the mean.

To check whether participants might have always reported less control in the together condition, agency ratings were analysed specifically in trials in which the marble crashed. Agency ratings were modelled by the social context, the outcome, and their interaction. When the marble crashed, results showed that only the outcome—how many points were lost—influenced agency ratings [$b = 2.28$, $t_{(25.07)} = 2.25$, $P = 0.034$, 95% CI = (0.39, 4.37)], with higher ratings associated with smaller losses. Social context no longer predicted agency ratings [$b = 0.36$, $t_{(25.57)} = 0.23$, $P = 0.82$, 95% CI = (−2.52, 3.55)], and there was no significant social context by outcome interaction [$b = 0.47$, $t_{(26.72)} = 0.30$, $P = 0.77$, 95% CI = (−2.66, 3.70)]. We further checked that according to the task design, outcomes did not differ, on average, across social contexts [Alone: mean = −85.06, SD = 2.92; Together: mean = −85.41, SD = 3.29; paired samples *t*-test: $t_{(26)} = 0.38$, $P = 0.71$]. Therefore, the relation between agency ratings and social context described earlier was specifically related to those trials in which the participant successfully acted.

To fully characterise participants' behaviour in the task, we also analysed number of trials in which the marble crashed, and in which the 'Other' agent acted instead (in the together condition). The marble crashed significantly more often in the alone condition (mean = 20.47%, SD = 8.63), than when playing together [mean = 15.00%, SD = 6.57; paired samples *t*-test: $t_{(26)} = 3.73$, $P < 0.001$]. In the together condition, the co-player acted significantly more often (mean = 19.44%, SD = 8.62) than the marble crashed [paired samples *t*-test: $t_{(26)} = 4.05$, $P < 0.001$]. These results, together with the earlier finding of later stops in the together condition, show that participants adapted their behaviour in order to minimise their losses in the together condition, when the "co-player" could act instead of the participant. To assess whether this strategy really was beneficial, we averaged the outcomes across all trials (successful stops, marble crashes and 'co-player' actions) for each participant. Results confirmed that, overall, participants lost significantly less points in the together condition (mean = −21.10, SD = 3.76), relative to playing alone [mean = −28.17, SD = 4.06; paired samples *t*-test: $t_{(26)} = -6.84$, $P < 0.001$]. Since the comparisons above showed no significant differences in outcomes across social contexts for successful stops, nor for marble crashes, this overall reduction in losses was clearly driven by the 'co-player' action trials, in which the participant did not lose any points.

ERPs

Mean amplitudes for the FRN component were analysed with the same model as agency ratings. Results revealed that FRN amplitude was significantly reduced (i.e. more positive) when playing together, relative to the alone condition [$b = 1.26$, $t_{(188.52)} = 2.40$, $P = 0.017$, 95% CI = (0.042, 2.28); see Figure 3]. FRN amplitude was not significantly influenced by the outcome [$b = 0.18$, $t_{(50.58)} = 0.37$, $P = 0.71$, 95% CI = (−0.83, 1.23)], nor by stop position [$b = -0.53$, $t_{(28.02)} = -1.00$, $P = 0.32$, 95% CI = [−1.56, 0.53)]. There were no significant interactions (see Supplementary Table S4).

Discussion

To investigate the cognitive and neural consequences of diffusion of responsibility, we developed a task in which participants either played alone, or together with another agent who could act instead of them. The best outcome for the participant occurred if they refrained from acting, but the co-player acted. The worst outcome occurred if neither participant acted. The co-player's presence led participants to act later, reduced their subjective sense of agency, and also attenuated the neural processing of action outcomes, as reflected by the FRN.

Behaviour

In the 'Together' condition, participants acted later and rated their feeling of control over action outcomes as lower, compared with 'Alone' trials. Importantly, participants had the same objective control over outcomes in 'Alone' and 'Together' trials. Further, the social context varied randomly between trials. Therefore, our results show that behavioural decisions and sense of agency are continuously updated by social context information.

In accordance with studies using implicit measures of agency (Takahata *et al.*, 2012; Yoshie and Haggard, 2013), we found that sense of agency was reduced for more negative outcomes. This shows that, as instructed, participants rated their

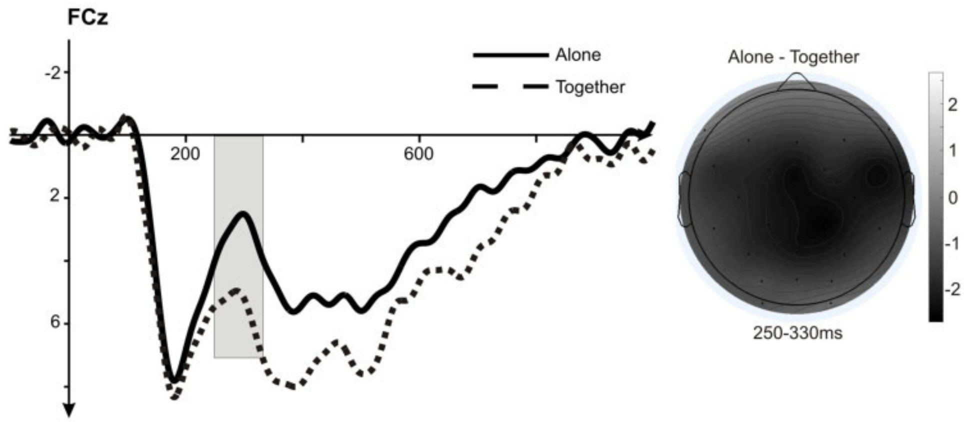

Fig. 3. ERPs. Grand average time courses are shown for the two experimental conditions. The analysed time window for the FRN (250–330 ms) is highlighted in grey. Topoplot shows the scalp distribution of the difference between the conditions averaged across the FRN time window.

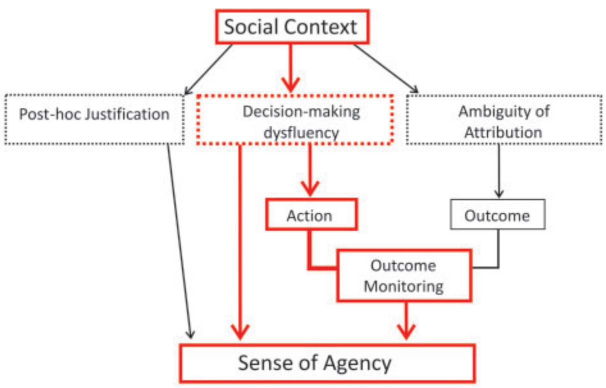

Fig. 4 The model shows different ways in which the presence of others may influence outcome monitoring and sense of agency. The pathways in black show mechanisms which can explain findings of previous studies, but are, as we show in this study, not necessary for diffusion of responsibility to occur. The central pathway (in red) shows the mechanism we propose, which can explain the observed effects in the absence of ambiguity and post-hoc justification.

subjective sense of control over the number of points they lost, rather than over whether the marble crashed. Reduced sense of agency over more negative outcomes could reflect the self-serving bias of attributing negative outcomes to external factors (Bandura, 1999). However, outcome magnitude effects in the 'Together' condition were no larger than in the 'Alone' condition, suggesting that social diffusion of responsibility does not simply reflect a misattribution of negative outcomes to others.

FRN

ERP results showed an effect of social context on the neural processing of action outcomes. In otherwise identical trials, FRN amplitude to outcomes of successful actions was reduced by the co-player's presence. Interestingly, we observed these effects on absolute amplitude of the FRN, rather than the commonly used difference wave between processing of positive and negative outcomes.

The FRN has been shown to be sensitive to distribution of control among multiple players (Li *et al.*, 2010), and to participants' beliefs regarding whether outcomes are controllable (Yeung *et al.*, 2005; Li *et al.*, 2011). In our task, instructions about action-outcome contingencies were identical between conditions, and full control remained with the participant. Thus, the mere presence of another player was sufficient to evoke changes in the neural processing of action outcomes akin to those observed when control over an outcome is abolished. As such, our EEG findings offer an objective measure consistent with subjective agency ratings. Attentional demands during the outcome processing were identical for 'Alone' and 'Together' trials. The FRN is thought to be sensitive to the motivational significance of outcomes (Gehring and Willoughby, 2002; Holroyd and Yeung, 2012). While in our task there was no 'objective' reduction in control over outcomes in 'Together' trials, participants nevertheless reported feeling less control over outcomes when the other player was present. Thus, the motivation to learn from such outcomes could be weakened, leading to reduced outcome monitoring.

Importantly, at the beginning of the outcome phase, participants knew they would lose a certain number of points, depending on where they stopped the marble. Therefore, participants' expectations could be assumed to be identical in Alone and Together trials. At the beginning of Together trials, participants may have anticipated the possibility of a better outcome (losing no points), than at the outcome of Alone trials. However, if this affected their outcome processing after they made an action, this should result in a larger FRN amplitude, as there would be a greater negative mismatch between anticipated and actual outcome.

Implications for concepts of diffusion of responsibility

Our findings significantly extend current models of diffusion of responsibility (Bandura, 1999), by demonstrating an online effect of social context on outcome processing. This is in line with Bandura's proposition that negative consequences of one's actions are less relevant in a group than in an individual context (Bandura, 1999). Social context might reduce the experience that actions are linked to their consequences. Bandura (1991) distinguishes diffused responsibility and distorted processing of action consequences as independent causes of reduced subjective responsibility. Our findings suggest that these phenomena could be related. Specifically, the presence of another agent can attenuate the processing of action outcomes, potentially leading to reduced sense of agency and responsibility. Consistently, coercion reduces sense of agency and attenuates the sensory processing of action outcomes (Caspar *et al.*, 2016).

The mechanisms underlying these effects remain unclear. Dysfluency in action selection has been shown to reduce sense of agency (Chambon *et al.*, 2014). Mentalising about potential behaviour of an alternative agent may increase the complexity of decision-making, and increase uncertainty about possible scenarios. We propose that such decision-making dysfluency may play a causal role in diffusion of responsibility as shown in Figure 4. Deciding whether to act is harder when someone else might also intervene, compared with when acting alone. We suggest that the need to mentalise in social contexts would complicate action selection, which in turn reduces sense of agency when one does act.

We designed our task to eliminate ambiguity in agency attribution, that is, about 'who' caused a given outcome. In real-life social situations, attribution of outcomes is likely to be more ambiguous than in comparable non-social situations. Reduced monitoring of action consequences due to the presence of others may then increase the likelihood of attribution errors. Correctly attributing consequences to their causes; however, is a critical prerequisite for learning action-outcome associations and likely also for forming a sense of moral responsibility. In this sense, the social dilution of agency might potentially have both immediate and longer-term effects on agency learning.

Conclusions

We show that diffusion of responsibility is not merely a *post-hoc* phenomenon reflecting a self-serving bias, but an online influence on how people process and experience the consequences of their actions. The presence of other agents can lead to reduced outcome monitoring and a reduction in individual sense of agency, even in the absence of attributional ambiguity. Distributed responsibility could potentially also cause differences in attentional demands or differences in outcome expectations. These effects could be mediated by increased complexity of decision-making processes in social situations.

Supplementary data

Supplementary data are available at SCAN online.

Funding

F.B. was supported by the German Research Foundation (BE 5779/1-1). N.S. was supported by a UCL Impact Scholarship, and by Belgian Science Policy Office project 'Mechanisms of conscious and unconscious learning' (IAP P7/33) to P.H. S.B. was supported by an AHRC Science in Culture Research Grant to P.H., and by the 'Subjectivity, Agency and Social Cognition' Project funded by the 'New Directions in the Study of the Mind' Project at Cambridge University. P.H. was supported by ERC Advanced Grant HUMVOL.

Conflict of interest. None declared.

References

Baayen, R.H., Davidson, D.J., Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59**(4), 390–412.

Bagiella, E., Sloan, R.P., Heitjan, D.F. (2000). Mixed-effects models in psychophysiology. *Psychophysiology*, **37**(1), 13–20.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, **50**(2), 248–87.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, **3**(3), 193–209.

Bandura, A., Underwood, B., Fromson, M.E. (1975). Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of Research in Personality* **9**(4), 253–69.

Bates, D., Maechler, M., Bolker, B., *et al.* (2014). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-7). Available: http://cran.r-project.org/web/packages/lme4/index.html

Bradley, G.L. (1995). Group influences upon preferences for personal protection: A simulation study. *Journal of Safety Research*, **26**(2), 99–105. http://doi.org/10.1016/0022-4375(95)00009-F

Caspar, E.A., Christensen, J.F., Cleeremans, A., Haggard, P. (2016). Coercion changes the sense of agency in the human brain. *Current Biology*, **26**(5), 585–92.

Chambon, V., Sidarus, N., Haggard, P. (2014). From action intentions to action effects: how does the sense of agency come about?. *Frontiers in Human Neuroscience*, **8**, 320.

Chekroun, P., and Brauer, M. (2002). The bystander effect and social control behavior: the effect of the presence of others on people's reactions to norm violations. *European Journal of Social Psychology*, **32(6)**, 853-67

Darley, J.M., Latane, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, **8(4p1)**, 377.

Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, **134**(1), 9–21.

Efron, B., Tibshirani, R.J. (1994). *An Introduction to the Bootstrap*. Boca Raton, Florida, USA: CRC Press.

Forsyth, D.R., Zyzniewski, L.E., Giammanco, C.A. (2002). Responsibility diffusion in cooperative collectives. *Personality and Social Psychology Bulletin*, **28**(1), 54–65.

Frith, C.D. (2014). Action, agency and responsibility. *Neuropsychologia*, **55**, 137–42.

Gehring, W.J., Willoughby, A.R. (2002). The Medial Frontal Cortex and the Rapid Processing of monetary gains and losses. *Science*, **295**(5563), 2279–82.

Hakstian, A.R., Suedfeld, P., Ballard, E.J., Rank, D.S. (1986). The ascription of responsibility questionnaire: development and empirical extensions. *Journal of Personality Assessment*, **50**(2), 229–47.

Holroyd, C.B., Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, **16**(2), 122–8.

Karau, S.J., Williams, K.D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, **65**(4), 681–706.

Kühn, S., Nenchev, I., Haggard, P., Brass, M., Gallinat, J., Voss, M. (2011). Whodunnit? Electrophysiological correlates of agency judgements. *plos One*, **6**(12), e28657.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2015). lmerTest: Tests in Linear Mixed Effects Models (Version 2.0-29). Available: https://cran.r-project.org/web/packages/lmerTest/index.html

Li, P., Han, C., Lei, Y., Holroyd, C.B., Li, H. (2011). Responsibility modulates neural mechanisms of outcome processing: An ERP study: modulation of outcome processing by responsibility. *Psychophysiology*, **48**(8), 1129–33.

Li, P., Jia, S., Feng, T., Liu, Q., Suo, T., Li, H. (2010). The influence of the diffusion of responsibility effect on outcome evaluations: Electrophysiological evidence from an ERP study. *NeuroImage*, **52**(4), 1727–33.

Lopez-Calderon, J., Luck, S.J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, **8**, 213.

Meier, B.P., Hinsz, V.B. (2004). A comparison of human aggression committed by groups and individuals: an interindividual–intergroup discontinuity. *Journal of Experimental Social Psychology*, **40**(4), 551–9.

Mynatt, C., Sherman, S.J. (1975). Responsibility attribution in groups and individuals: a direct test of the diffusion of responsibility hypothesis. *Journal of Personality and Social Psychology*, **32**(6), 1111.

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available: http://www.R-project.org

Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, **80**(1), 1.

San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience*, **6**, 304.

Schel, M.A., Brass, M., Haggard, P., Ridderinkhof, K.R., Crone, E.A. (2014). Neural correlates of intentional and stimulus-driven inhibition: a comparison. *Frontiers in Human Neuroscience*, **8**, 27.

Takahata, K., Takahashi, H., Maeda, T., *et al*. (2012). It's Not My Fault: Postdictive Modulation of Intentional Binding by Monetary Gains and Losses. *PLoS One*, **7**(12),e53421.

Tibon, R., Levy, D.A. (2015). Striking a balance: analyzing unbalanced event-related potential data. *Quantitative Psychology and Measurement*, **6**, 555.

Timm, J., SanMiguel, I., Keil, J., Schröger, E., Schönwiesner, M. (2014). Motor intention determines sensory attenuation of brain responses to self-initiated sounds. *Journal of Cognitive Neuroscience*, **26**(7), 1481–9.

Wallach, M.A., Kogan, N., Bem, D.J. (1964). Diffusion of responsibility and level of risk taking in groups. *The Journal of Abnormal and Social Psychology*, **68**, (3), 263.

Yeung, N., Holroyd, C.B., Cohen, J.D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex*, **15**(5), 535–44.

Yoshie, M., Haggard, P. (2013). Negative emotional outcomes attenuate sense of agency over voluntary actions. *Current Biology*, **23**(20), 2028–32.