Shared responsibility decreases the sense of agency in the human brain

Marwa El Zein[1,2]*, Ray J.Dolan[3,4] and Bahador Bahrami[2,5,6]

[1] Centre for Adaptive Rationality, Max Planck Centre for Human Development, Berlin, Germany

[2] Institute of Cognitive Neuroscience, University College London, London, United Kingdom

[3] Max Planck University College London Centre for Computational Psychiatry and Ageing Research, University College London, London, UK

[4] Wellcome Centre for Human Neuroimaging, University College London, London, UK

[5] Faculty of Psychology and Educational Sciences, Ludwig Maximilian University, Munich, Germany

[6] Department of Psychology, Royal Holloway, University of London, Egham, Surrey, United Kingdom

*Corresponding author: Marwa El Zein marwaelzein@gmail.com

Abstract

Distributing responsibility across several individuals in social decisions helps minimize a burden of responsibility for consequences of decisions. Here we investigated the neural expression of this effect using magnetoencephalography (MEG). Participants performed a reward-based learning decision-making task in contexts where their sense of responsibility over outcomes decreased with group size. An MEG outcome processing effect was reduced as a function of decreasing responsibility at 200ms post outcome onset and centred over parietal and precentral brain regions. During social decisions, prior to outcome revelation, a motor preparation signature at 500ms after stimulus onset was attenuated. A boost in responsibility for positive outcomes in social contexts was associated with increased activity in regions related to social and reward processing. Together, these results show that sharing responsibility with others reduces agency through an influence on pre-outcome motor preparation and post-outcome processing, affording an opportunity for flexibility in credit for positive outcomes.

Introduction

The brain mechanisms underlying responsibility and executive control for our actions have generally been addressed by comparing active vs forced or passive choices choice[1–4]. The subjective experience of a coerced action, where one is instructed what to do, renders a feeling of diminished responsibility that is akin to a passive action and is associated with reduced neural processing of an action's outcome[1]. This perspective is germane to isolated individuals who are free to choose and who are solely responsible for ensuing consequences, or conversely when a decision is made on their behalf not being accountable at all. However, in social contexts, for example when voting, we find ourselves somewhere in between the above extremes.

In collective choice, we have less control than when deciding alone, though not completely constrained by being instructed what to do. In return for this partial concession of control to a collective, we benefit from a *sharing* of responsibility for our choices[5], while still allowing us to flexibly claim credit when outcomes are good. Thus, when people assign credit to contributors in a team, they tend to overestimate their own contribution and overattribute a team's success to their own ability, effort and merit[6]. Perhaps more alarmingly, collective context allows us to disavow ourselves of regret[7] as well as offload blame onto others[8] when outcomes are not good. Correspondingly, people find it harder to punish groups (vs individuals) who have violated a social norm[8]. In 2003, when weapons of mass destruction were not found in Iraq, or indeed over the following years, intelligence agencies whose reports had justified a catastrophic invasion of Iraq defended themselves by invoking that "*everyone had agreed at the time*".

Previous neural investigations of responsibility have had a focus on the evaluation of outcomes[1,9–11]. Examining brain responses to decision outcomes has afforded a very convenient methodological handle on the complicated concept of responsibility. However, this has detracted from a more nuanced understanding of the neural basis of this complex construct, where a sense of responsibility does not emerge with knowledge of an outcome. A clear empirical demonstration of the extended nature of responsibility encompassing deliberation, choice, action and outcome is the Bystander Effect[12]. When someone needs our help but our engagement is expected to be costly, the presence of other bystanders reduces the stress and anticipated blame for not helping. The Bystander Effect illustrates that we experience the burden of responsibility even before we commit to a choice to help or not and,

importantly for our purpose here, this burden is modulated by social context. The neural underpinnings of this deliberative aspect of responsibility are unknown.

Here, we investigate two unresolved facets of a sense of responsibility, where we operationally define responsibility as a participant's subjective judgements about the causal attribution between a decision and outcome. In this, we follow an earlier literature that proposes responsibility is tightly related to the 'sense of agency' and feeling of control[1,5,13,14]. In a gambling choice task, participants made decisions that led to positive or negative outcomes, while responsibility was parametrically modulated through the impact of different social contexts. We constructed four different contexts: (1) *private* in which the individual participant assumed full responsibility; (2) *dyadic* and (3) *group* in which the individual decided together with 1 or 4 other people, respectively and shared the responsibility with them, and (4) *forced* where another person decided on behalf of the participant, exonerating all responsibility. This experimental setup allowed us to assess responsibility in a parametric (rather than categorical) design. Previous studies show that performing actions together with others reduces subjective ratings of responsibility and control[10,15,16]. Based on these prior findings, we expected responsibility to be highest in the *private* context and progressively decrease in *dyadic* and *group,* and at its lowest in the *forced* context (**Hypothesis 1a**). Critically, the action-outcome association frequencies were kept constant across all conditions.

We used magnetoencephalography (MEG), which provides a high temporal resolution neural signal, to disentangle the dynamics in neural processes underlying responsibility at various stages of the task. Our parametric design allowed us to conduct trial by trial regressions of responsibility contexts on MEG signals[17,18], instead of categorically comparing grand-average event-related fields under high vs low responsibility as previously done[1,9–11]. Our neuronal hypotheses are formulated based on a prediction that the neural correlates of responsibility are common to those the support a sense of agency. Previous studies found a decrease in the neural signatures of outcome processing resulting with reduced responsibility, for example when people are coerced to perform an action (vs willingly perform the same action), an auditory tone signalling outcome of the action evokes a lower auditory evoked potential (N1)[1]. Moreover, the feedback related negativity (FRN) - an evoked potential that appears around 200 to 300ms after outcome onset – is also attenuated when an outcome resulted from an action performed in the presence of another agent[9], during an uncontrollable task[11], and with others[10]. Other research on the neural basis of a sense of agency shows that a fronto-parietal brain network, including the inferior parietal lobule, the angular gyrus and the premotor and motor cortices, are implicated in subjective feeling of control[2,19–21]. Based on

these findings, we predicted an attenuated outcome processing as a function of decreased responsibility, in particular at 200-300 ms after outcome onset with localisation to a fronto-parietal network (**Hypothesis 2**).

To identify neural signatures of outcome-independent deliberative responsibility, we examined a sense of agency MEG signal in the time-period prior to outcome. It has been suggested that when deliberating an action, our sense of agency arises from a mental simulation of that action[14,22,23] and deliberative responsibility, in this view, consists of mental simulation of likely outcomes of those imagined actions. One study, where participants underwent functional Magnetic Resonance Imagery (fMRI) while reading vignette scenarios and imagining themselves or others as the protagonist, showed contemplation of the consequences of imagined actions recruited premotor cortex[24] consistent with a mental simulation account. A fronto-parietal brain network implicated in the sense of agency includes premotor and motor cortices, and is known to be involved in action intention[2,19–21]. Thus, we predicted that prior to outcome, preparation of motor activity would be modulated by responsibility context. Specifically, we hypothesized that lateralized motor preparation signals[25] would be attenuated in social contexts where responsibility is reduced (**Hypothesis 3**).

Finally, we addressed an influence of outcome valence and its interaction with responsibility. Previous research has shown people tend to follow a self-serving bias, whereby they take more credit for positive outcomes as compared to negative outcomes, an effect evident in both group and individual decisions[26–30]. On this basis we expected to observe increased responsibility ratings for positive vs negative outcomes irrespective of whether participants decided alone or with others, but not when they had no responsibility for the decision, i.e., in the *forced* context (**Hypothesis 1b**). In addition, we hypothesized that, since social contexts offer a possibility of sharing responsibility, this would enable participants to cherry pick credit for positive outcomes and offload blame for negative outcomes on to others[5], thereby exaggerating a self-serving bias in the two social vs Private context. Importantly, our experimental design allowed us to examine two distinct possible mechanisms: a concurrent increase of credit and decrease in blame or specific modulation of either the credit or the blame respectively (**Hypothesis 1c**).

Results

Behavioural results

Participants performed a task where they had to choose between two gambles that could yield to positive or negative outcomes. They did so in four different contexts: Alone (Context **Private**), with one other player (Context **Dyadic**), in a group of five players (Context **Group**), and another player deciding for them (Context **Forced**) (**Figure 1a**). In this final Forced context, participants were requested to press the right or left button even though it did not count as their choice. As the button they had to press was known, participants were faster to respond in a Forced context compared to all the other contexts (all T_{39}>6.4, all p<0.001). Participants were also faster to make a decision in the Dyadic as compared to Group context (T_{39}=-2.63, p=0.01, μ=-15.94, ci=[-28.18, -3.71], dz=0.41), but no other differences were observed. After the decision between the two gambles, participants reported on a scale how responsible they felt for the outcome of the trial. These responsibility ratings represent our main behavioural variable of interest (**Figure 1b**).

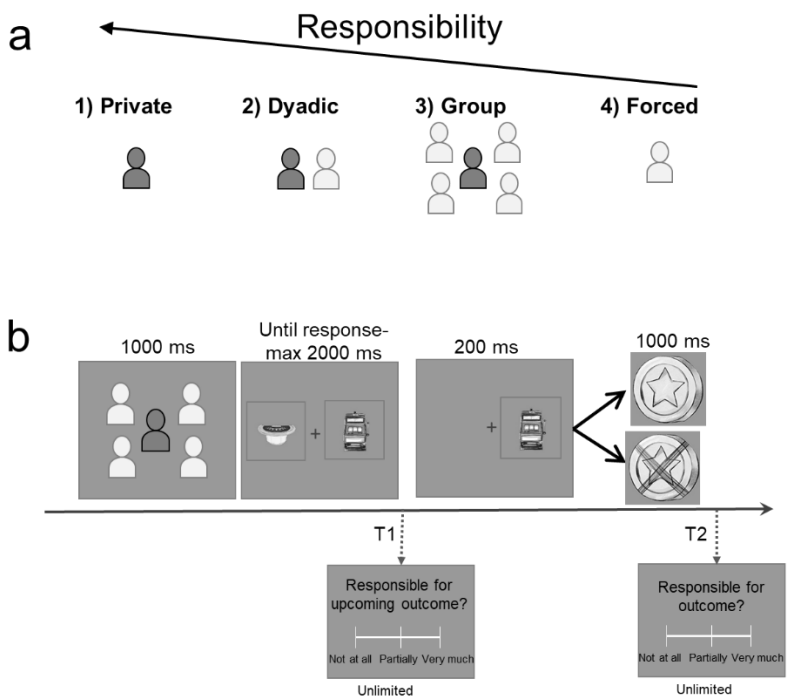

Figure 1. Experimental design. a) Participants performed a decision-making task in four contexts: (1) Private: playing alone, (2) Dyadic: together with another participant, (3) Group: together within a group of five participants, and (4) Forced: another participant played for them. Dyadic and Group contexts are referred to as Social context. b) Each trial (384 trials in 8 blocks), began with participants seeing, for 1 second, the context (as described in *a*) in which they would be playing. Next, they chose between two gamble options displayed on the screen for a maximum of 2 seconds. After choice, the selected gamble remained on the screen for 200 ms. In the Social context, the selected gamble may or may not coincide

with the chosen gamble. Finally, the outcome (positive or negative) was displayed for 1 second. In randomly assigned one third of the trials of each block, immediately after gamble selection, here marked by T1, participants rated how responsible they felt for the upcoming outcome. In another third of the trials, after outcome (T2), participants rated how responsible they felt for the obtained outcome. In the remaining one third of the trials no rating was required.

Parametric responsibility reports

To test our first **hypothesis 1a**, for each participant, we regressed z-scored responsibility ratings against responsibility contexts (1= Private, 2= Dyadic, 3= Group, 4= Forced)**.** All participants except 1 ($p>0.2$) showed a significant negative slope (37 participants $p<0.001$, 1 participant $p<0.005$, 1 participant $p<0.02$), i.e., linearly reported being more responsible from the Private to the Dyadic to the Group to the Forced context (T-test of the betas computed for each participant against zero $T_{39}=-19.72$, $p<0.001$, $\mu=-0.61$, ci=[-0.67,-0.55], dz=-3.11) (**Figure 2a**). Group level t-test comparisons between contexts confirmed this linear change in responsibility: Responsibility ratings were higher in the Private as compared to the Dyadic context ($T_{39}=12.14$, $p<0.001$, $\mu=0.76$, ci=[0.63,0.89], dz=1.95). Dyadic context ratings were higher than Group context ratings ($T_{39}=6.08$, $p<0.001$, $\mu=0.24$, ci=[0.16,0.33], dz=0.96). Finally, Group context ratings were higher than the Forced context ($T_{39}=13.19$, $p<0.001$, $\mu=0.96$, ci=[0.81,1.10], dz=2.08).

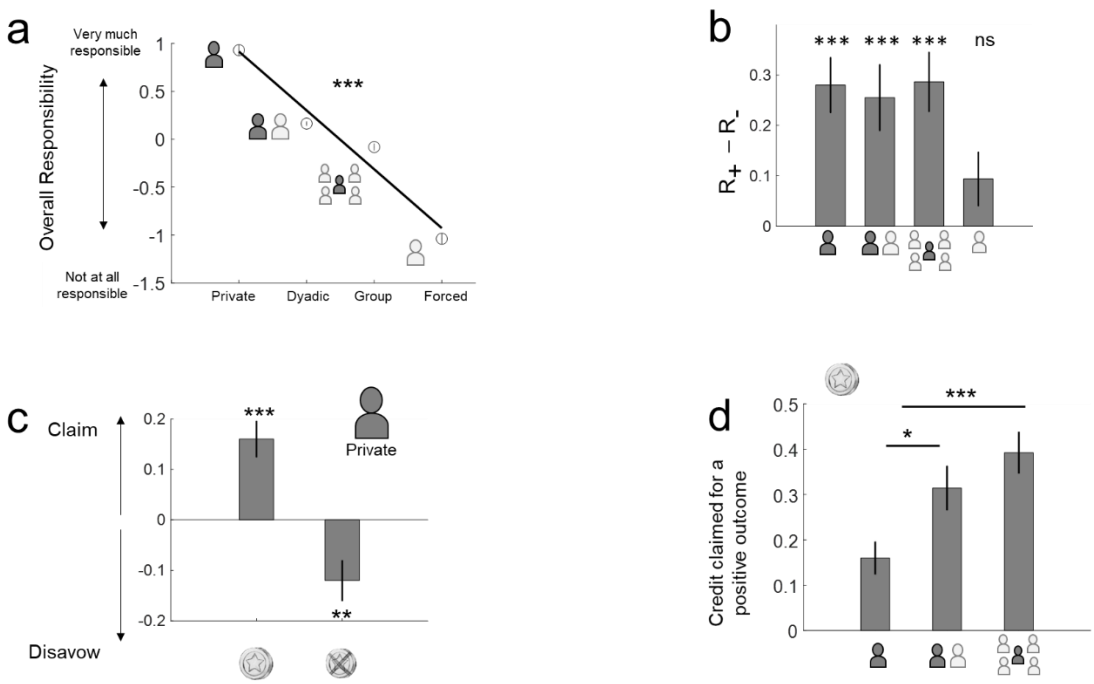

Figure 2. Behavioural results. Responsibility ratings were reported on a continuous scale from 'not at all responsible' to 'very much responsible' (see Fig 1.b). The ratings were z-scored for each participant for analyses and are therefore reported in arbitrary units in all graphs. a) Parametric decrease of reported responsibility from Private to Dyadic to Group to Forced contexts. Circles represent the mean values across participants for each context with the associated standard errors. The line is the mean

general linear model fit (fit within each participant then meaned across all 40 participants). b) The impact of outcome valence on reported responsibility. Y axis shows the difference in responsibility claimed for positive (R_+) and negative (R_-) outcomes. In all cases where the participant had *some* choice in the selection of the gamble, they claimed more responsibility for positive (vs negative) outcomes. The Forced context, where the subject had no say in gamble selection did not show a similar bias. c) When compared to reported responsibility *before* outcome, we observed outcome valence modulated responsibility in both positive and negative directions. Data from Private context alone: participants both claimed more credit after positive outcomes and disavowed a negative outcome. Y axis shows the difference between responsibility ratings after vs before outcome. X axis shows outcome valence. d) People claimed more credit for a positive outcome in the social context when their vote was selected as compared to the private context. Error bars represent standard errors. ***: $p<0.001$, **: $p<0.01$, *$p<0.05$, *ns*: non significant.

These behavioural results confirm that responsibility is reduced in social contexts, and further show this reduction depends on group size. Moreover, a consistent parametric decrease in responsibility reports validates our experimental paradigm designed to address the neural processes involved in (social) decision-making under different responsibility levels.

Self-serving bias

Our behavioural **hypothesis 1b** was that responsibility ratings would reveal a self-serving bias, with participants attributing more responsibility to themselves for positive (vs negative) outcomes. We observed participants indeed provided higher responsibility ratings for positive (vs negative) outcomes in the 3 active contexts (Private context: $T_{39}=5.18$, $p<0.001$, $\mu=0.28$, ci=[0.17,0.39], dz=0.82; Dyadic context: $T_{39}=3.92$, $p<0.001$, $\mu=0.25$, ci=[0.12,0.39], dz=0.61; and Group context: $T_{39}=4.89$, $p<0.001$, $\mu=0.29$, ci=[0.16, 0.40], dz=0.77), but not in the Forced context ($T_{39}=1.76$, $p>0.08$, $\mu=0.09$, ci=[-0.01,0.20], dz=0.27) (**Figure 2b**). The magnitude of this bias did not differ across the 3 active contexts (all $p>0.48$, all $T>0.69$). Contrary to our **Hypothesis 1c**, we found no evidence for an increase in self-serving bias in the social (vs Private) contexts.

Claiming or disavowing outcomes?

Next, we examined if an observed self-serving bias was due to attributing credit after positive outcomes or disavowing responsibility after negative outcomes. In different trials, we asked participants to report their responsibility ratings *before* and *after* outcome. Taking the *before* ratings as a baseline, we then assessed whether the self-serving bias consisted of an elevation of ratings after positive and/or a reduction of ratings after negative outcomes. For each outcome valence, we subtracted the *before* from *after* ratings.

In the Private context, a self-serving bias comprised both components: more responsibility was claimed after (vs before) positive outcomes ($T_{39}=4.41$, $p<0.001$, $\mu=0.16$, ci=[0.09,0.23],

dz=0.72) and less responsibility was accepted after (vs before) negative outcomes (T_{39}=-2.96, p<0.01, μ =-0.12, ci=[-0.20,0.03], dz=-0.46) (**Figure 2c**).

In social contexts, a more complex analysis was required to take account of the relationship between a participant's decision, the collective choice and outcome. When the collective decision matched the participant's decision, responsibility ratings were boosted after (vs before) outcome (T_{39}>4.26, p<0.001). Importantly here, participants claimed more credit (vs the Private context) for positive outcomes (Private vs Dyadic: T_{39}=-2.51, p<0.02, μ =-0.15, ci=[-0.27,-0.03], dz=-0.39; Private vs Group T_{39}=-4.39, p<0.001, μ =-0.23, ci=[-0.33,-0.12], dz=-0.70) (**Figure 2d**). This finding is partly consistent with **Hypothesis 1c** in revealing social contexts offered a 'cover' to claim more credit for a positive outcome. Participants however did not disavow responsibility for negative outcomes (T_{39}>0.59, p<0.55) which led to an opposite effect to our prediction: participants disavowed negative outcomes more in a private vs social contexts (Private vs Dyadic: T_{39}=-3.05, p=0.004, μ =-0.15, ci=[-0.25,-0.05], dz=-0.48; Private vs Group T_{39}=-2.79, p=0.007, μ =-0.14, ci=[-0.25,-0.04], dz=-0.42). In trials where the social choice was different from the participant's choice, responsibility ratings were generally reduced after (vs before) an outcome (T_{39}>5.3, p<0.001) (see supplementary material for a full description of these results).

MEG results

Outcome processing is parametrically modulated by shared responsibility

Our key neural hypothesis was that neural signatures of shared responsibility would be common to those of a sense of agency. **Hypothesis 2** was that outcome processing within 200-300ms after outcome onset would vary linearly with responsibility levels and this signal would locate to a fronto-parietal brain network associated with a sense of agency that includes inferior parietal lobule, angular gyrus and premotor and motor cortices. We isolated pre-processed MEG signals in this time window and performed whole-brain regressions of those signals against the responsibility contexts (see Methods). Two clusters of electrodes showed significant activity which survived correction for multiple comparisons (cluster alpha < 0.05, **Figure 3a** right panel): these electrodes match an expected fronto-parietal network in consisting predominantly of central, frontal, and parietal electrodes (see full list of electrodes in supplementary materials). The mean parameter estimate of the regression on the right sided significant electrodes is shown through time in figure 3a left panel (statistics of the effect at the peak in the 200-300ms window at 220ms (T_{39}=-4.61, p<0.001, μ =-4.15, ci=[-5.98,-2.33], d=0.72). As an additional descriptive analysis, on these same right electrodes, we computed

event related fields locked to outcome, and separately for each context: the largest event-related field was observed for the Private context and decreased linearly until the Forced context at ~200 ms after outcome onset (**Figure 3b**), in line with the trial-by-trial GLM results. Finally, the same GLM regressions in the source space (see Methods) revealed that the parametric encoding of responsibility at 200-300ms after outcome onset is associated with parietal and central sources (**Figure 3c**) described in detail in Table 1.

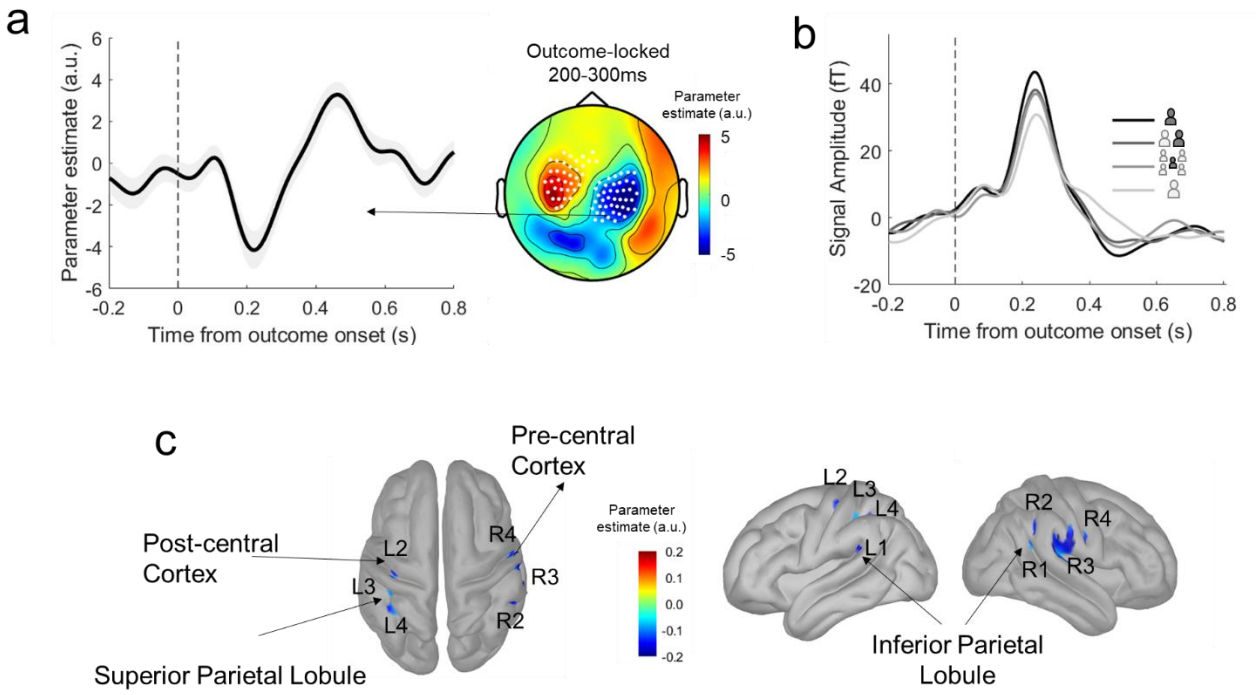

Figure 3. Agency-related neural correlates of responsibility at 200 to 300ms after outcome onset. a) Parameter estimate of responsibility regression. Right panel: Scalp topography of the parameter estimate of responsibility regression [1 Private, 2 Dyadic, 3 Group 5, 4 Forced] on the mean MEG activity at 200-300ms after outcome. White dots represent the significant electrodes where MEG signal linearly co-varies with the level of responsibility using cluster corrections of the effect against zero at an alpha cluster level <0.05. Left panel: Parameter estimate of responsibility regression at the right cluster locked to outcome onset. b) Associated Event-related Fields (ERF) at the same right electrodes cluster showing how the amplitude of the ERF locked to outcome increases with responsibility. c) Estimated cortical sources of the responsibility parameter estimate at 200-300ms. Parameter estimates of the responsibility regression at the source level that are significant at p<0.01 are shown. . R=Right hemisphere and L=Left Hemisphere, and the associated numbers refer to the different brain regions reported and anatomically localized in Table 1.

[Insert Table 1]

Neural correlates of *deliberative* (outcome-independent) sense of responsibility

Having established a neural signature of socially shared responsibility in outcome processing, we then investigated a neuronal expression of deliberative responsibility under private and social contexts. Recall that by deliberative, we refer to responsibility experienced prior to

choice and outcome, which we hypothesised would be related to motor preparation activity and subject to modulation by responsibility context. Specifically, we tested whether lateralized motor preparation signals to select a gamble with the left or right hand[25] are decreased in the social contexts where responsibility is reduced. We first computed motor preparation signals at 100 ms before response button press by subtracting the power in the 8-32Hz frequency band when participants responded with the right hand from when they responded with the left hand which allowed to identify the central electrodes with maximal suppression (**Figure 4a** top panel). Then, we subtracted the power in the 8-32Hz frequency band in ipsilateral minus contralateral maximal central electrodes relative to the hand pressed (**Figure 4a** bottom panel). We found that motor preparation signals increase gradually until response in all four contexts. Note a motor preparation signal in the Forced context, where the button press was known and the reaction times were fastest, diverged radically from the other conditions and these were excluded from further hypothesis testing (see Supplementary materials).

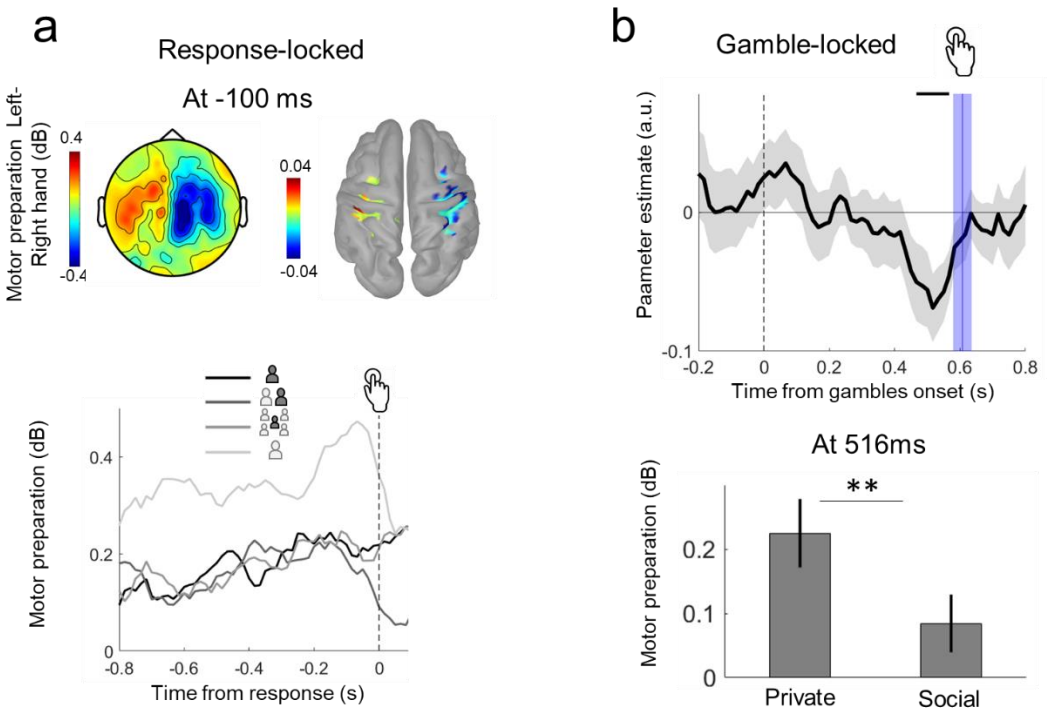

Figure 4. Motor preparation signals are modulated by responsibility context. a) Topography showing the mean power of 8-32Hz frequencies at 100ms before the motor response for conditions where participants answered with the Left Hand minus the Right Hand and the associated estimated sources. Bottom panel: Response-locked motor lateralization: motor preparation from 0.8 second before until response measured with the power of 8-32Hz Frequencies in ipsilateral minus contralateral electrodes relative to the hand pressed locked to the motor response in all four different contexts. b) Stimulus-locked motor lateralization: Top panel: Parameter estimate of the regression of motor preparation signal against the Social vs Private regressor. Negative parameter estimates indicate lower motor preparation in the Social vs Private context, which is significant around 500 ms after gambles onset. The black line indicates timepoints where the parameter estimate is significant against zero (cluster one-tailed pcorr<0.05). The blue patch indicates the time where the button was pressed based on reaction times and their standard errors. Bottom panel: motor preparation signal at the peak of the effect at 516ms for the private vs the social context. ** p<0.01

To test our **Hypothesis 3**, we asked whether motor preparation signals locked to gamble onset differed in the Private vs Social context. The parameter estimate of this regression was significant at ~500ms after gamble stimulus onset (Peak of parameter estimate at 516 ms T_{39}=-2.81, p=0.007, μ =-0.068, ci=[-0.11, -0.02], d=0.44, cluster from 466 ms to 566 ms one-tailed pcorr<0.05, **Figure 4b** top), where a stronger motor preparation signal was evident for Private compared to Social contexts (Direct t-test at the peak 516ms between Private and Social Contexts T_{39}=2.78, p=0.008, μ=0.14, ci=[0.03,0.24], dz=0.45; **Figure 4b** bottom). No significant cluster was observed for the same analyses locked to motor response rather than gamble onset (p>0.1).

Neural correlates of the boost in credit claim in social contexts

Earlier we provided behavioural analyses that showed a social vs private setting boosted the credit people claimed for positive outcomes. In an exploratory analysis we examined the neural correlates of this specific positive boost, focusing on Private and Social contexts for trials where the collective choice matched the participants' vote. Concentrating on a post-outcome 200-300ms time window, we identified MEG signals locked to *positive* outcomes and then ran a GLM with Social vs Private as regressor. This revealed a significant cluster (cluster alpha <0.05) that included frontal, temporal and central electrodes (see full list of electrodes in supplementary materials, **Figure 5a**). No significant clusters appeared for the same regression analysis for MEG signals locked to *negative* outcomes. The same regression conducted at MEG source level for signals locked to *positive* outcomes revealed source estimates in the Orbito-frontal cortex and temporal lobe, including the Superior Temporal Sulcus (see details of brain regions in Table 2, **Figure 5b**). In the pre-central cortex, a negative parameter estimate was observed reflecting a higher activity in the Private context in comparison to the Social context, consistent with the effects of responsibility observed in this region.

[Insert Table 2]

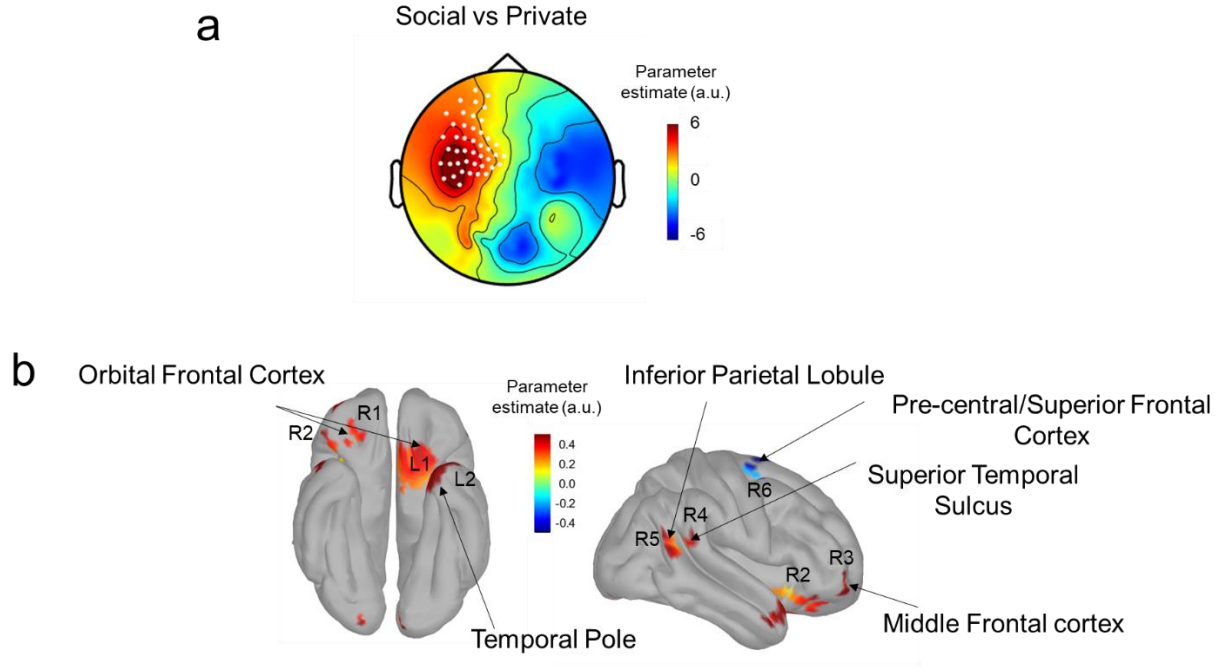

Figure 5. The processing of positive outcomes in Social vs Private context. a) Scalp topography of the parameter estimate of the regression Social vs Private on the mean MEG activity at 200-300ms after positive outcomes following decisions that matched the participant's vote. White dots represent the significant cluster of electrodes differentiating between Social and Private outcome processing using cluster corrections of the effect against zero (cluster alpha <0.05). b) Estimated cortical sources of the Social vs Private parameter estimate at 200-300ms. Parameter estimates that are significant at p<0.01 are shown. R=Right hemisphere and L=Left Hemisphere, and the associated numbers refer to the different brain regions reported and anatomically localized in Table 2.

Discussion

We examined behavioural and neuronal signatures of shared responsibility in collective judgments. Behaviourally, we show responsibility in collective decision making was reduced compared to private individual decision making, and this reduction depended on the collective size. More importantly, our key hypothesis that collective context attenuates responsibility by reducing a sense of agency provided the basis for testable neurobiological predictions. Previous neurobiological findings on social responsibility have consistently shown a decreased outcome processing under coercion[1], cooperative gambling[10], and in the presence of another person[9]. Our work extends on these studies in several important respects. First, our study examines neuronal signatures of responsibility when participants are engaged in a collective choice, in the context of a systematic variation in group size that facilitated a parametric design, with responsibility level progressively reduced across levels. Finally, the previous neural studies on responsibility have, to date, exclusively focused on outcome evaluation[1,4,9–11]. Here, we examined the neural effects related to the sense of responsibility when deliberation and action preparation take place before a choice or outcome are known.

People's subjective reports of responsibility varied with social context, with greater responsibility when deciding privately compared to with others, in line with previous studies showing that people feel less control, responsibility and regret when acting with others[7,10,15,16]. Responsibility reports varied with group size, with more responsibility in dyads vs in a group of five. This finding extends previous results showing that in problem solving people take less credit when they are in a larger group[31].

The increased responsibility ratings for positive vs negative outcomes (Fig. 2b) replicates the "self-serving bias" effect whereby participants take more credit for positive (vs negative) outcomes[26,29,30]. Our experimental design included a subset of trials where we probed responsibility *before* outcome was declared. This allowed us to treat these trials as baseline and evaluate if the observed self-serving bias was due to claiming more credit for positive outcomes and/or offloading blame for the negative ones. In addition, our design also distinguished between trials in which the group choice was in line with that of the participant or not. Consequently, collective choices in which the group and participant agreed offered a particularly informative situation wherein an individual retained *some* control but still shared the responsibility with others. Under these situations, participants were particularly disposed to claim disproportionately more credit for positive outcomes. These findings point to a potential motivational lure of joining groups, particularly as claiming credit for success has been shown to increase self-esteem[29]. Interestingly, in these trials, we did not observe any

offloading of blame onto others for negative outcomes. This latter observation is consistent with a similar recent report[32].

We show that at 200-300ms after outcome, MEG recorded activity over bilateral fronto-parietal brain regions decreased linearly, from highest under full responsibility to shared responsibility in dyads, to shared responsibility in groups, to no responsibility in the forced context. This complements previous findings of decreased outcome processing associated with low responsibility contexts[1,9,10], but goes beyond those earlier works by showing a parametric -- and not categorical – relationship to a stepwise manipulation of responsibility. Second, using source estimation we confirmed our key neural hypothesis that a marker of responsibility is localized to brain regions previously associated with a sense agency in pre-central and post-central cortices, and superior and inferior parietal lobules[2,19–21,33]. Our findings are also consistent with previous work that identified correlates of motor intentions in the parietal cortex[2] and motor control in inferior parietal lobule[20].

It is important to note that an outcome-dependent neural signature of agency identified here (Fig. 3) was not confounded by outcome valence. In our design we were mindful to ensure a probability of winning did not depend on gamble choice. Winning probability was also independent of whether a participant decided privately, did not decide at all or when the participant's choice was approved at a social level or not. Furthermore, to minimize incidental learning, we used a large set of visual stimuli (i.e., 40 illustrations of various gamble devices) and had the participants choose between randomly sampled pairs.

Several non-monetary motivational factors came into play in deciding with others in our paradigm. The first factor was autonomy and control. Rewards have a higher salience when we are instrumental in obtaining them compared to when they have just happened to be thrust upon us. Our key neural findings (Fig. 3) are consistent with this by showing that the participant's level of involvement in an outcome modulates an outcome-dependent neural signal. Thus, brain responses to outcomes were stronger when subjects decided privately and parametrically decreased as responsibility decreased. The second factor was approval from others. Previous work shows that others' approval is, inherently, capable of driving the brain's motivational reward network even when no monetary reward is at stake[34]. In our paradigm, participants' choices were sometimes approved, and other times overruled by the collective. Accordingly, we found approval of the collective was associated with increased neural activity the Orbital Frontal Cortex (OFC), the Superior Temporal Sulcus and the Temporal pole, brain regions variously associated with reward, social processing and mentalizing[35,36]. In addition to being associated with monetary reward and value processing[37,38], the human OFC is implicated in individual differences in conformity and reaction to other people's opinion[39]. As the processing of an outcome resulting from a social decision involves considering other

people's responsibility, the involvement of the Superior Temporal Sulcus (STS) is in line with the finding that this regions is involved when participants consider others' responsibility[24].

Previous studies of the neural basis of responsibility have invariably focused on decision and action outcome. Our study breaks with this tradition and examines neural substrates of responsibility prior to outcome, when deliberation and action preparation are taking place. We show that the motor preparation (to pick the right or left gamble) around 500ms after onset of a visual display of a gamble option was reduced under shared (social) as compared to full responsibility (private) conditions. These results were observed only when the analysis was stimulus locked, but not locked to the motor response, indicating that the decreased motor preparation in social vs private contexts most likely reflects a deliberative rather than motor process. This is also in line with findings that self-responsibility as compared to other-responsibility recruits brain areas associated with action simulation including the premotor context, suggesting that higher order social processes may relate to simple goal-directed action[24]. A recent theoretical model proposed that a decreased sense of agency in social contexts relates to mentalizing processes as people need to take into account others perspective[9,32]. This model posits that through mentalizing, social contexts increase decision dysfluency and action planning. Here we provide the neural evidence that action planning is indeed reduced when people make decisions with others in social contexts.

Our novel experimental design manipulating levels of responsibility over decision outcomes shows that responsibility over similar outcomes influences their processing at 200ms in brain regions that are related to a sense of agency. Our results also provide for the first time an outcome-independent neural signature of responsibility evident in a reduction in pre-outcome motor preparation signatures at 500ms in shared responsibility contexts, i.e., social contexts. The finding that deliberative and retrospective responsibility share neural mechanisms with a sense of agency potentially advances an understanding of the complex notion of societal responsibility and is relevant to a wide range of societal domains including legal, medical, and artificial intelligence-related ethical issues.

Methods

Participants

46 healthy adults (24 females, mean age=24.13±4.42) participated in a magnetoencephalography (MEG) study, with 2 participants excluded due to technical errors in saving the data or in the triggers' information. 4 participants were excluded due to noisy MEG data noted at the pre-processing phase (their data had more than 10 noisy channels and/or more than 15% of trials removed after visual inspection). This left a total of 40

participants (21 females, age=24.00±4.46). The study was conducted at the Wellcome Centre for Neuroimaging, 12 Queen square, London WC1B 5JS. Participants were recruited by email via the University College London (UCL) SONA platform and the Institute of Cognitive Neuroscience (ICN) participants' pool. All participants were aged 18-36 years, right-handed, with a normal or corrected-to-normal vision and had no neurological or psychiatric history. They provided written informed consent according to the regulations approved by the UCL research ethics committee (Project ID Number 9923/002). Participants were informed that they will receive £25 for their participation and a bonus up to £5 based on their gains in the experiment. All participants were accorded the bonus and compensated £30.

Experimental design

Stimuli were generated using Cogent 2000 and Cogent Graphics toolboxes running in MATLAB (MathWorks, Na-tick, MA, USA). The task was an apparent learning game where participants had to choose between 2 different gambles that were supposed to have different probabilities of getting a positive or negative outcomes. The 2 gambles corresponded to 2 different images of hand-drawn realistic gambling devices among 40 total drawings (an example of 2 drawings is shown in **Figure 1b**). 2 gambles were pseudo-randomly presented at each trial among a set of 10 gambles at each block: we controlled for the gambles to be presented approximately the same amount of time within each context (contexts described below, **Figure 1a**) and each block, and not more than 3 times in a raw. Unbeknownst to participants, the gambles were not associated with different probabilities yielding to positive or negative outcomes but did so at chance level.

In each trial, following a fixation cross displayed for 700-900ms, participants first saw a cue (duration 1000ms) indicating which context they would be playing in. There were four possible contexts (**Figure 1a**): (1) **Private**, (2) **Dyadic:** with one other player such as both players make a decision but only one of their decisions is selected to determine the outcome, (3) **Group:** with 4 other people where the picked gamble is based on a majority vote procedure (the gamble picked by 3 or more people) – (2) and (3) are referred to as **Social** context – and (4) **Forced**: another player chose on behalf of the participant. After the cue, a fixation cross was displayed for a variable period of 1000-1200ms. Then, two gambles appeared on the screen each on the left and right side of the cross, and participants had to select between them by pressing the buttons on the devices they held with their right (Right gamble) and left (Left gamble) hands (**Figure 1b**).

In the Forced context, participants were instructed to press the same (either the left or the right hand) button after seeing the gambles (choice of which button to press was counterbalanced across participants). This button press was included to ensure that action

requirements for all trial types were identical, but not indicate any choice and allowed us to keep identical stimulus-response mappings. Participants were told that different gambles had different probabilities of winning and losing and that they should try and pick the one that had a higher chance of winning. In all trials, the response window was 2 seconds, otherwise the trial was classified as missed (even in the Forced context). The selected gamble was then displayed for 200ms followed by a variable blank period of 650-850ms. The trial ended with an outcome (positive/negative) presentation (1000ms). Participants were told that one trial would subsequently be picked from each block and that they would earn a cumulative bonus based on whether the outcome on selected trials was positive or negative, with a missed trial counting as a negative outcome. Note that positive outcomes allowed them to win a bonus, while negative outcomes were similar to 'neutral' outcomes as they could not 'lose' any money.

In one third of the trials, immediately after having made a decision but *before* an outcome (i.e., T1 see **Figure 1b**), participants rated how responsible they felt for an upcoming outcome (from 'not at all' to 'partially' to 'very much' on a continuous scale). In another third of the trials, they made a similar rating immediately *after* the outcome (i.e., T2). In one third of the trials, no ratings were required.

Participants completed 384 trials in 8 blocks of 48 trials each. Each block was composed of a balanced number of trials: 2 contexts (Alone, Forced) X 2 Outcomes (Positive/Negative) X 3 Scales (T1, T2 or none) X 2 repetitions, and in the social contexts: 2 contexts (Dyadic, Group) X 2 Outcomes (Positive/Negative) X 3 Scales (T1, T2 or none) X 2 Feedback (whether the participant's vote was selected or not).

Behavioural analyses

Responsibility ratings, collected on a continuous scale, were z-scored for each participant before subject and group-level statistical analyses and are reported in arbitrary units. A general linear model (GLM) was performed for each participant's responsibility ratings with the parametric responsibility contexts as a regressor [1,2,3,4]: 1= Private, 2= Dyadic, 3= Group, 4= Forced**.** The betas of these regressions were tested against zero with a t-test for the group level statistics. Paired student t-tests were used to compare ratings in each context from one another, to assess the differences between ratings for positive and negative outcomes, and differences between ratings both before after an outcome. Reaction times for the gamble decision were also compared between contexts using paired student t-tests. Confidence intervals of the effects and effect size (cohen's d) are provided.

MEG acquisition and pre-processing

MEG data was recorded using a 275-channel CTF Omega system whole-head gradiometer (VSM MedTech, Coquitlam, BC, Canada). Neuromagnetic signal was continuously recorded at a 600 Hz sampling rate. After participants were comfortably seated in the MEG, head localizer coils were attached to the nasion and 1 cm anterior of the left and right tragus to monitor their head movement during recording. Due to technical issues, three gradiometers were disabled from the system: MLO42, MRC12, MLT54, leaving a total of 272 instead of 275 channels. Three additional channels recorded eye movements (x, y movement and pupil diameter) using an eye-tracker (SR Research Eyelink non-ferrous Infra-red eye tracking system).

Stimuli were projected at a 60Hz frequency on a screen of size 42 x 32 cm, with a distance between screen and eyes of about 60 cm. During piloting, a photodiode placed on the screen allowed us to measure a delay between the trigger signal and the projection of stimuli of about 33 ms. The appearance of stimuli on the screen was therefore monitored with a photodiode attached at the lower edge of the screen. A black square was presented there for 50 ms at the time where the stimulus was presented. The signal from the photodiodes was recorded in parallel to the other MEG channels to provide a precise temporal marker for the appearance of stimuli. All triggers were re-matched to the signal of the photodiode associated with each stimulus of interest.

We used Fieldtrip[40] to process the data. The raw MEG data was notched for the 50, 100 and 150Hz power line noise before visual inspection (combined with an automatic detection of artifacts) to reject trials with jumps and strong muscular activity. The data was low-pass filtered at 35Hz and epoched using the photodiode signal locked to context onset, gamble onset, response onset and outcome onset. Independent component analysis (ICA) was performed on the data to reject eye blinks, saccades and heartbeat components thanks to the visual inspection of the components. The mean proportion of rejected artifacted trials was 7.7%± 3.1 for context-locked data, 6.5%±3.4 for gamble locked data, 6.4% ± 3.4 for response-locked data and 5.9%±2.8 for outcome-locked data.

Time-frequency decompositions were performed by computing the spectral power of the 8-32Hz frequency bands using multitapering transform (Slepian tapers, 8-32Hz, four cycles) centred on context presentation, gamble presentation and response onset, using the fieldtrip function ft_freqanalysis. Power spectrum was extracted for the main analysis, and the complex Fourier spectrum was extracted for the source reconstruction analyses.

MEG analyses

Regression analyses of outcome-locked MEG signals

We performed single-trial regressions of MEG signals against variables of interest:

- Responsibility context [1,2,3,4] as for the behavioural analyses: 1= Private, 2= Dyadic, 3= Group, 4= Forced. Note that negative parameter estimates indicate higher activity for more responsibility.
- Social (+1 for the Dyadic and Group contexts) vs Private (-1) contexts.

As our hypothesis predicted that outcome processing would be modulated by responsibility at 200-300ms, we performed regressions on the mean activity of MEG signals at 200-300ms after outcome onset at each of the 272 electrodes. Beta coefficients of regressions within each participant were tested against zero for the group level statistics. Multiple comparisons corrections were performed in the electrode space using ft_prepare_neighbours in fieldtrip, coupled with non-parametric Montecarlo statistics to determine the clusters of electrodes where these regressions were significant with a p-value<0.05 (cluster corrections, clusteralpha=0.05, test statistic set as the maximum level of the cluster-level statistic, alpha=0.05, 10000 randomizations, 2 minimal neighbouring channels).

Motor preparation measures in the time-frequency domain

Motor lateralization

The suppression of 8-32Hz frequency bands in the hemisphere contralateral to the hand used in a motor press response provides a neural marker for motor preparation[25]. For each subject, the power of this frequency band when participants responded with their right hand was subtracted from the power when they responded with their left hand, at 100ms before a response press. When averaged across subjects, this allowed us to pick the central electrodes with maximal suppression (**Figure 4a**). Motor lateralization for responses with the right and left hand was obtained by subtracting power activity in the central electrodes contralateral to the utilised hand from power activity in the central electrodes ipsilateral to the hand used, thus resulting in a positive motor preparation as shown in **Figure 4b** (top panel).

Right electrodes: 'MRC13'; 'MRC14'; 'MRC22'; 'MRC31'; 'MRC32'; 'MRC41'; 'MRC42'; 'MRC53'; 'MRC54'; 'MRC55'.

Left electrodes: 'MLC13'; 'MLC14'; 'MLC22'; 'MLC31'; 'MLC22'; 'MLC41'; 'MLC42'; 'MLC53'; 'MLC54'; 'MLC55'.

We performed regressions on this motor lateralization measure at each time point, locked to the gamble and to the response using the following regressor:

- Social context (+1 pooling both Dyadic and Group) vs Private (-1)

We added reaction times (z-scored) as an additional regressor, to account for effects over and above differences in reaction times for motor press. The parameter estimates were tested for significance against zero at each time point from 0 to 600ms after gamble onset with multiple comparison corrections across time points implemented using non-parametric cluster-level statistics [41]. The pairing between experimental conditions and MEG signals was shuffled 1,000 times, and the maximal cluster-level statistics (the sum of one-tailed t-values across contiguously significant time points at a threshold level of 0.05) extracted for each shuffle to compute a 'null' distribution of effect size across a time window of [0,+600] ms locked to stimulus presentation, or [-600,0] locked to response onset. For significant clusters in the original (non-shuffled) data, we computed the proportion of clusters in the null distribution whose statistics exceeded the one obtained for the cluster in question, corresponding to its 'cluster-corrected' p-value.

Source reconstruction

Minimum-norm source estimates were performed using Brainstorm [42]. We computed Kernel inversion matrices for each subject, and each of the 8 blocs (recommended because of participants' head movements from a bloc to another), based on all trials consisting of non-overlapping time windows locked to context, gamble and outcome. A generic brain model was used taken from the default anatomy in Brainstorm: ICBM 152 Nonlinear atlas version 2009 [43,44]. The head model was computed with an overlapping spheres method. The noise co-variance was computed based on 900 ms baseline before the context onset (i.e, the baseline of the whole trial). Sources were computed with minimum norm imaging and the current density map method and following recommendations when no individual anatomical IRM is available, we chose unconstrained solutions to source estimation. In unconstrained source maps, there are three dipoles with orthogonal orientations at each cortex location (15002 vertices X 3 orientations = 45006 X 272 electrodes, 8 Inversion matrices –for each of the 8 blocks - for each participant). To display them as an activity map and perform statistics, the norm of the vectorial sum of the three orientations at each vertex is computed:

$$S = \sqrt{S_x^2 + S_y^2 + S_z^2}$$

GLM in the source space

For each subject, at 200-300 ms after outcome onset, the mean MEG signals were multiplied by individual Kernel matrices for each block, and the norm was computed before implementing the GLMs at each vertex at the source level. Finally, the betas at each vertex were averaged

across blocks resulting in one map of parameter estimates for the effect of interest for each participant. Then, for group level analyses, t-tests against zero were done at each vertex and only the vertices with mean parameter estimates significant across participants at p-value<0.01 are shown. To assess the anatomical location of the significant vertices, we first report regions based on the Destrieux Atlas[45] that is provided in Brainstorm. Second, we extract MNI coordinates of regions and project them onto The Human Brainnetome Atlas BNA[46] (through MRICron where MNI coordinates can be matched to the anatomical location of chosen Atlas NITRC: MRIcron: Tool/Resource Info).

Motor lateralization

For each participant, at 100 ms before response press, the MEG fourier transforms from the time-frequency decomposition analysis were multiplied by individual Kernel matrices for each frequency and trial. Then, after taking the power of the magnitude of the complex fourier spectrum, a mean was performed on the frequencies (8-32Hz) and the conditions where the participant responded with the Right Hand and the conditions where the participant responded with the Left Hand. The sum of the power for each orientation was computed for each block, before doing a mean on the 8 blocks to obtain one activity map per condition and per subject. Left press vs Right press were then contrasted by performing a t-test in the source level after a z-score of activity map per participant, and keeping only significant vertices at p<0.01 for the figure. The mean difference across all participants at the vertices with this significant effect is shown (**Figure 4a** bottom).

References

1. Caspar, E. A., Christensen, J. F., Cleeremans, A. & Haggard, P. Coercion Changes the Sense of Agency in the Human Brain. *Curr. Biol.* **0**, (2016).
2. Desmurget, M. *et al.* Movement Intention After Parietal Cortex Stimulation in Humans. *Science* **324**, 811–813 (2009).
3. Haggard, P., Clark, S. & Kalogeras, J. Voluntary action and conscious awareness. *Nat. Neurosci.* **5**, 382–385 (2002).
4. Kool, W., Getz, S. J. & Botvinick, M. M. Neural representation of reward probability: evidence from the illusion of control. *J. Cogn. Neurosci.* **25**, 852–861 (2013).
5. El Zein, M., Bahrami, B. & Hertwig, R. Shared responsibility in collective decisions. *Nat. Hum. Behav.* **3**, 554–559 (2019).
6. Farwell, L. & Wohlwend-Lloyd, R. Narcissistic Processes: Optimistic Expectations, Favorable Self-Evaluations, and Self-Enhancing Attributions. *J. Pers.* **66**, 65–83 (1998).
7. El Zein, M. & Bahrami, B. Joining a group diverts regret and responsibility away from the individual. *Proc. R. Soc. B Biol. Sci.* **287**, 20192251 (2020).
8. El Zein, M., Seikus, C., De-Wit, L. & Bahrami, B. Punishing the individual or the group for norm violation. *Wellcome Open Res.* **4**, 139 (2019).
9. Beyer, F., Sidarus, N., Bonicalzi, S. & Haggard, P. Beyond self-serving bias: diffusion of responsibility reduces sense of agency and outcome monitoring. *Soc. Cogn. Affect. Neurosci.* **12**, 138–145 (2017).
10. Li, P. *et al.* The influence of the diffusion of responsibility effect on outcome evaluations: electrophysiological evidence from an ERP study. *NeuroImage* **52**, 1727–1733 (2010).
11. Li, P., Han, C., Lei, Y., Holroyd, C. B. & Li, H. Responsibility modulates neural mechanisms of outcome processing: an ERP study. *Psychophysiology* **48**, 1129–1133 (2011).
12. Darley, J. M. & Latane, B. Bystander intervention in emergencies: Diffusion of responsibility. *J. Pers. Soc. Psychol.* **8**, 377–383 (1968).

13. Frith, C. D. Action, agency and responsibility. *Neuropsychologia* **55**, 137–142 (2014).

14. Haggard, P. Sense of agency in the human brain. *Nat. Rev. Neurosci.* **18**, 196–207 (2017).

15. Dewey, J. A., Pacherie, E. & Knoblich, G. The phenomenology of controlling a moving object with another person. *Cognition* **132**, 383–397 (2014).

16. Nicolle, A., Bach, D. R., Frith, C. & Dolan, R. J. Amygdala involvement in self-blame regret. *Soc. Neurosci.* **6**, 178–189 (2011).

17. El Zein, M., Wyart, V. & Grèzes, J. Anxiety dissociates the adaptive functions of sensory and motor response enhancements to social threats. *eLife* **4**, (2015).

18. Wyart, V., de Gardelle, V., Scholl, J. & Summerfield, C. Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron* **76**, 847–858 (2012).

19. Chambon, V., Wenke, D., Fleming, S. M., Prinz, W. & Haggard, P. An online neural substrate for sense of agency. *Cereb. Cortex N. Y. N 1991* **23**, 1031–1037 (2013).

20. Desmurget, M. & Sirigu, A. Conscious motor intention emerges in the inferior parietal lobule. *Curr. Opin. Neurobiol.* **22**, 1004–1011 (2012).

21. Haggard, P. Neuroscience. The sources of human volition. *Science* **324**, 731–733 (2009).

22. Gallagher, S. Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* **4**, 14–21 (2000).

23. Jeannerod, M. Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage* **14**, S103-109 (2001).

24. Blackwood, N. J. *et al.* Self-responsibility and the self-serving bias: an fMRI investigation of causal attributions. *NeuroImage* **20**, 1076–1085 (2003).

25. Donner, T. H., Siegel, M., Fries, P. & Engel, A. K. Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Curr. Biol. CB* **19**, 1581–1585 (2009).

26. Caine, B. T. & Schlenker, B. R. Role Position and Group Performance as Determinants of Egotistical Perceptions in Cooperative Groups. *J. Psychol.* **101**, 149–156 (1979).

27. Forsyth, D. R. & Schlenker, B. R. Attributing the causes of group performance: Effects of performance quality, task importance, and future testing. *J. Pers.* **45**, 220–236 (1977).

28. Jaquiery, M. & El Zein, M. Stage 1 Registered Report: How responsibility attributions to self and others relate to outcome ownership in group decisions. *Wellcome Open Res.* **6**, 24 (2021).

29. Leary, M. R. & Forsyth, D. R. Attributions of responsibility for collective endeavors. in *Group processes* 167–188 (Sage Publications, Inc, 1987).

30. Yoshie, M. & Haggard, P. Negative emotional outcomes attenuate sense of agency over voluntary actions. *Curr. Biol. CB* **23**, 2028–2032 (2013).

31. Forsyth, D. R., Zyzniewski, L. E. & Giammanco, C. A. Responsibility Diffusion in Cooperative Collectives. *Pers. Soc. Psychol. Bull.* **28**, 54–65 (2002).

32. Sidarus, N., Travers, E., Haggard, P. & Beyer, F. How social contexts affect cognition: Mentalizing interferes with sense of agency during voluntary action. *J. Exp. Soc. Psychol.* **89**, 103994 (2020).

33. Chaminade, T. & Decety, J. Leader or follower? Involvement of the inferior parietal lobule in agency. *NeuroReport* **13**, 1975–1978 (2002).

34. Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J. & Frith, C. D. How the opinion of others affects our valuation of objects. *Curr. Biol. CB* **20**, 1165–1170 (2010).

35. Frith, C. D. & Frith, U. Social Cognition in Humans. *Curr. Biol.* **17**, R724–R732 (2007).

36. Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* **42**, 9–34 (2014).

37. Lopez-Persem, A. *et al.* Four core properties of the human brain valuation system demonstrated in intracranial signals. *Nat. Neurosci.* **23**, 664–675 (2020).

38. Padoa-Schioppa, C. & Conen, K. E. Orbitofrontal Cortex: A Neural Circuit for Economic Decisions. *Neuron* **96**, 736–754 (2017).

39. Campbell-Meiklejohn, D. K. *et al.* Structure of orbitofrontal cortex predicts social influence. *Curr. Biol.* **22**, R123–R124 (2012).

40. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
41. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
42. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* **2011**, 879716 (2011).
43. Fonov, V. *et al.* Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage* **54**, 313–327 (2011).
44. Fonov, V., Evans, A., McKinstry, R., Almli, C. & Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* **47**, S102 (2009).
45. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* **53**, 1–15 (2010).
46. Fan, L. *et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb. Cortex N. Y. NY* **26**, 3508–3526 (2016).

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (PI: BB grant agreement No. 819040 - acronym: rid-O) and from the Wellcome Trust (PI: RJD grant number 098362/Z/12/Z). MEZ was funded by the Wellcome Trust grant number 204702 and is currently supported by the European Union's Horizon 2020 Marie Curie Individual fellowship grant number 882936 acronym SIND. BB was supported by the Humboldt Foundation and by the NOMIS foundation. RJD is in receipt of a Wellcome Trust Investigator Award.

We would like to thank Daniel Bates for his valuable help with data acquisition, the MEG team at the Wellcome Trust Center for neuroimaging for their feedback on the project, David Wurzer for his help with data pre-processing and Alizee Lopez for her help with source analyses.

Author contributions

M.E.Z and B.B designed the study. M.E.Z programmed the experiment, collected and analysed the data. M.E.Z, R.D and B.B discussed analyses and interpreted results. M.E.Z and B.B wrote the paper and R.D provided important revisions.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and material requests should be addressed to Marwa El Zein marwaelzein@gmail.com

Tables

	Destrieux Atlas			MNI coordinates			Human Brainnetome Atlas (BNA)			
Regions	Index	Name	Localisation	X	Y	Z	Label ID	Gyrus	Anatomy	Lobe
Left 1 (L1)	41	Posterior segment of the lateral sulcus	Insula	-48	-40	25	145	Inferior parietal lobule	Rostroventral area 40	Parietal
Left 2 (L2)	29	Precentral gyrus	Lateral frontal lobe	-41	-24	57	155	Post central Gyrus	Area 1/2/3	
Left 3 (L3)	67	Postcentral Sulcus	Lateral Parietal Lobe	-34	-36	42	129	Superior Parietal Lobule	Lateral area 5	
Left 4 (L4)	56	Intraparietal sulcus and transverse parietal sulci		-34	-45	39	129			
Right 1 (R1)	73	Superior temporal sulcus	Temporal lobe	46	-44	24	144	Inferior parietal lobule	Rostroventral area 39	
Right 2 (R2)	55	Sulcus intermedius primus (of Jensen)	Inferior parietal lobule	53	-43	35	142		Caudal area 40	
Right 3 (R3)	26	Supramarginal Gyrus		58	-19	17	146		Rostroventral area 40	
	67	Post Central Sulcus	Main sulci of the lateral aspect of the parietal lobe	54	-15	35	160	Post central Gyrus	Area 2	
	36	Temporal plane of the superior temporal gyrus	Superior aspect of the temporal lobe	65	-26	16	76	Superior temporal Gyrus	Caudal area 22	Temporal
Right 4 (R4)	45	Central Sulcus	Frontal Lobe	51	-5	25	54	Precentral Gyrus	Area 4	Frontal

Table 1. Anatomical sources of parametric responsibility encoding locked to outcome. Regions were determined based on the Destrieux Atlas implemented in Brainstorm, then MNI coordinates coordinates were extracted from Brainstorm and projected on the BNA atlas.

	Destrieux Atlas			MNI coordinates			Human Brainnetome Atlas (BNA)			
Regions	Index	Name	Localisation	X	Y	Z	Label ID	Gyrus	Anatomy	Lobe
Left 1 (L1)	24	Orbital Gyri	Ventral aspect of the frontal Lobe	-21	14	-27				
	63	Medial orbital sulcus	Ventral aspect of the frontal Lobe	-14	17	-15	49	Orbital Gyrus	Area 13	Frontal
		Gyrus Rectus	Medial aspect of the frontal Lobe	-3	11	-23				
		Subcallosal Gyrus	Limbic Gyrus	-4	6	16				
Left 2 (L2)		Temporal pole	Superior aspect of the temporal lobe	-22	13	-43				
Left 3 (L3)		Triangular part of the inferior frontal gyrus	Main frontal gyri	-51	45	5				
		Inferior Frontal sulcus		-40	44	2.0	21	Middle Frontal Gyrus	Ventral area 9/46	Frontal
	15	Middle Frontal Gyrus		-49	49	3				
Left 4 (L4)	28	Postcentral Gyrus	Lateral aspect of the parietal lobe	-24	-31	77	161	Postcentral Gyrus	Area 1/2/3	Parietal
Left 5 (L5)	27	Superior parietal lobule	Superior parietal lobule	-37	-52	66				
Right 1 (R1)	64	Orbital Sulcus	Ventral aspect of the frontal Lobe	23	39	-22	46	Orbital Gyrus	Lateral area 11	Frontal
Right 2 (R2)		Orbital Gyri		43	25	-18	44	Orbital Gyrus	Orbital area 12/47	
Right 3 (R3)		Orbital Gyri		43	57	-11				
	15	Middle Frontal Gyrus	Main frontal gyri	39	63	-1	28	Middle Frontal Gyrus	Lateral area 10	Frontal
Right 4 (R4)	72	Superior temporal sulcus	Lateral aspect of temporal lobe	67	-46	15	144	Inferior parietal Lobule	Rostroventral area 39	Parietal
Right 5 (R5)	38	Middle temporal Gyrus		64	-54	15	144			
	25	Angular gyrus	Inferior parietal lobule	63	-56	19	144			
Right 6	69	Superior part of the precentral sulcus	Main frontal sulci	30	-4.5	64	8	Superior frontal gyrus	Dorsolateral area	Frontal

Table 2. Anatomical sources of Social vs Private processing of positive outcomes. Regions were determined based on the Destrieux Atlas implemented in Brainstorm, then MNI coordinates were extracted from Brainstorm and projected on the BNA atlas. Grey areas represent MNI coordinates that could not be matched on the BNA atlas.