

fMRI coordination game study

Limbic and prefrontal activity during conformity and violation of norms in a coordination game

Tim Hodgson, Francesco Guala, Tim Miller and Ian Summers

Key words: Society, emotion, cooperation, rewards, social groups

Abstract

Previous neuroimaging studies of scenarios such as the prisoner's dilemma and ultimatum game show how brain systems which evolved to process rewarding and aversive stimuli are activated differentially during adherence to and breaches of norms. One theory of how shared social norms evolve is that they arise from mutually rewarding conventions which, via repeated execution, acquire a normative value which sustains social cohesion even when individual self interest is not served. Here we report an fMRI study of a two player coordination game in which players must coordinate on an arbitrary convention (a left or right button press) to obtain monetary rewards. Once this convention has been established, one of the participants is incentivised to deviate from the equilibrium with the offer of extra reward. During the period prior to decisions to violate the convention, activity was observed in regions associated with reward processing, such as the mid-brain, caudate and orbitofrontal cortex. Activations in advance of decisions to continue coordinating included the amygdala and anterior insula / inferior frontal gyrus. **The data are discussed in the light of theories which propose the existence of multiple interacting value-based decision making systems in the human brain.**

INTRODUCTION

While economic and evolutionary analyses of social norms are well developed (Ullmann-Margalit, 1977; Elster, 1989; Bicchieri, 2006; Gintis, 2009) research into the neural basis of norm-driven behaviour is still in its infancy (Montague and Lohrenz 2007). To date these studies take the existence of norms as given and observe participants while they cope with dilemmas to either adhere to existing norms such as altruism or act in their own material self-interest (Rilling et al., 2002; Sanfey et al., 2003; deQuervain et al., 2004; Spitzer et al., 2007; Haruno and Frith, 2010). These designs are useful, but also lack realism. The subjects for example typically do not have a history of mutual cooperation before they face the experimental dilemma to violate (or conform to) a social norm. This contrasts with the well-established fact in evolutionary biology and anthropology that social norms tend to arise (and to be enforced) primarily in the context of repeated small-group interactions (Henrich and Henrich, 2007). Here we report an fMRI study of a repeated coordination game designed to model how social norms might evolve by repeated execution of an arbitrary but mutually rewarding behavioural convention. We then created dilemma situations during which one of the participants was given an incentive to deviate from the established convention.

The idea that society is a gigantic coordination game and that social norms emerge via the repeated play of such games goes back at least to the philosopher David Hume (Hume 1748). The idea has been revived and articulated in the last century by game theorists and social scientists interested in the emergence of cooperation (Schelling, 1960; Lewis 1969, Sugden 1986; Skyrms, 1996, 2004; Binmore, 1998, 2005). Empirical studies have shown that humans perform better than would be expected by chance at coordination games in the laboratory. One explanation for this is that people engage in 2nd order theorising about how others perceive the salience of competing choices in such tasks (Mehta et al., 1994). Behavioural studies also suggest that mutually rewarding conventions develop social normative value such that participants continue with a coordinating strategy even when their material interest is not directly served by it (Guala and Mittone, 2010). This tendency to comply with established norms may have evolved from a natural aversion to violating others' expectations and a tendency to seek others' approval (Sugden 1986, 1998). This is consistent with previous studies that have observed activity in limbic areas of the brain related to approach / avoidance behaviour and emotion during norm violation (Rilling et al., 2002; Sanfey et al., 2003).

The present study used fMRI to study brain activity in a two player coordination game. Once a coordinating convention / equilibrium had been established in the two players, one of them was given an incentive to deviate from the convention with the offer of additional monetary reward. We reasoned that similar patterns of activity should be observed in this situation as has been reported in previous fMRI studies of social/moral dilemmas. Based on these studies we predicted activation within regions associated with negative affect such as the anterior insula cortex and amygdala during norm violation (Hampton et al., 2007; Sanfey et al., 2003; Phillips et al., 1997; Adolphs et al., 2004). We also expected activations in structures implicated in the processing of rewards such as the mid-brain, striatum and orbitofrontal cortex during decisions that had social as well as monetary value to participants (Schultz, 2000; Rilling et al., 2002; Klucharev et al., 2009; O'Doherty et al., 2001; O'Doherty et al., 2002). Activity in other brain areas was expected when participants engaged in inhibition of reward driven responding or 2nd order reasoning about another's intentions. Previous studies have linked these processes to areas of the prefrontal cortex (Greene et al., 2004; Hare et al., 2009; Spitzer et al., 2007). In particular the rostral medial prefrontal cortex is an area which has been strongly implicated in tasks that demand mentalising about others (Hampton et al., 2008; Krueger et al., 2009) and formulating predictions about future actions.

METHODS

Behavioural task

The behavioural task comprised a coordination game in which both participants could gain monetary rewards by selecting either a left or right button press response. On each round participants were presented with a decision screen (Figure 1b) which depicted the payoff matrix for that round. After a decision had been made and following a delay period in which the screen was blank (see fMRI Methods and Analysis section below) an outcome screen was displayed (Figure 1c) describing what had happened on the last round including the choice of the other player, the gains made in that round and the cumulative gains made in the game thus far. Participants were told that the experiment would last for "approximately 20 rounds" although in fact the game always lasted 21 rounds (this was done so that players could not predict exactly when the game would end whilst at the same time being aware of its

approximate duration). On “Normal” rounds both players were rewarded for selecting the same button press. As well as these Normal coordinating rounds, participants were instructed beforehand that a number of “Special” rounds could also occur at unspecified points in the game. They were told that on these Special rounds the payoffs may change and some players may not be told about the changes. The actual monetary payoff structure used in the game for Normal and Special rounds is shown in Figure 1a. On these rounds the potential deviant participant, hereafter referred to as “Player 1”, was given a monetary incentive to deviate from the established coordinating convention, whilst “Player 2” was still only rewarded for continued coordination. For both rounds types Player 1 always saw both their own and others payoffs whilst Player 2 only had information concerning their own payoffs in the task. Unknown to the participants, Special rounds were always programmed to occur on rounds 9, 13, 17 and 21. The instruction sheet included an example payoff matrix of the type that they would later see on the screen during the experiment on Normal rounds.

Following the final round, Player 2 was presented with a final decision screen which gave them the option to punish Player 1 for deviating on Special rounds. Both participants were unaware that this punishment decision would occur at the end of the experiment as the instructions given prior to the experiment did not inform them that this would happen. The amount of monetary punishment for Player 1 was equivalent to the payout received by player 1 on Special rounds for which they opted to deviate (i.e. £2 x number of deviations). However, the instruction screen also informed Player 2 that if they **chose** to deduct this amount from Player 2 their total winnings would be deducted by a smaller sum equivalent to a Normal coordinating round payout (50p) x the number of Special rounds on which Player 1 choose to deviate. Following Player 2’s decision Player 1 was shown an outcome screen which indicated whether or not Player 2 had opted to punish them by deducting money.

At the end of the testing session participants were given the money that they had won in the game, typically between £5 and £18 per participant depending on the outcome of the game and whether they had been assigned as Player 1 or Player 2.

The experimental task was implemented using custom software developed within the Z-tree programming environment for experimental economics (Fischbacher, 2007). The non-scanned participant performed the task on a computer in the scanner control room linked to the

fMRI coordination game study

computer used by the scanned participant via a local Ethernet connection. Visual stimuli were presented to the scanned subjects on a projection screen which they viewed via a mirror mounted on the head coil as they lay prone in the scanner. Choice decisions for the scanned participants were recorded using a press button box connected to a computer in the control room via a fibre optic link. The non-scanned participants indicated their choices using a mouse click to an on-screen button.

Participants

32 pairs of participants took part in the study. All scanned participants were recruited from the student population (graduate and undergraduate) of the University of Exeter (UK). All participants gave their informed consent to participate in the study and the procedures were approved by the School of Psychology ethics committee. In 20 of the sessions Player 1 was the scanned participant and in 12 sessions Player 2 was scanned. For 8 out of the 32 participant pairs, Player 2 was in fact a member of the MR centre research team and Player 1 was the scanned participant. In these cases, where Player 2 was one of the research team, the other player did not knowingly meet them, but was told that there was another player already seated in the scanner control room ready to start the experiment. For the other participant pairs, both Players filled out consent forms and read the task instructions whilst seated in the same waiting room before the scanning session, but were not permitted to converse with each other before or during the test. Of the scanned participants 13 were male and 19 female (mean age: 25 range: 19-48) whilst of the volunteer non-scanned participants, 12 were male and 12 female (mean age: 24 ranging from 19 to 36).

FMRI Methods and analysis procedure

The experiment was carried out using a 1.5T Philips Gyroscan Intera Scanner in the Peninsula MR Research Centre, St Lukes campus, University of Exeter. A T2*-weighted echoplanar sequence was used (TR = 3000ms, TE = 50ms, flip angle 90°, 32 transverse slices, 3.6 x 3.6 x 4mm, ascending acquisition). 125 volumes were acquired in each of the 4 runs per subject. An additional 5 “dummy” scans were performed at the start of each block prior to the start of the stimulus sequence. Data were analysed using SPM5 software (www.fil.ion.ucl.ac.uk/spm). The images were realigned, unwarped to remove variance caused

fMRI coordination game study

by movement-by-field inhomogeneity interactions, normalised to a standard EPI template, and smoothed with a Gaussian kernel of 6mm full-width at half maximum.

On each round the BOLD response was assessed for two “mini-epochs”: (1) the Decision period during which the participant was viewing the payoff matrix for that round and deciding which button to press; and (2) the “Outcome period” during which he / she was viewing the choices made by self other on that round and the monetary payments made to each player. The duration of the Decision period epoch was determined by the participants’ reaction time to press one of the two buttons from stimulus onset up to a maximum of 20 seconds. As soon as the participants had made their decision the display screen was blanked out until the display of the Outcome period screen, which was always presented 30 seconds after the onset of the decision period screen. The duration of the Outcome period was constant at 10 seconds (corresponding to the time the outcome display was presented on the screen). A further 12 seconds then elapsed prior to the onset of the next trial’s Decision screen.

Each participant’s fMRI data was analysed using a statistical model comprising a series of statistical regressors derived by convolving task event onsets and durations with a canonical hemodynamic response function. The GLM approach was used to assess correlations between the measured BOLD response and each regressor of interest. For the first level (individual subject) analysis translational and rotational head movement parameters were included as nuisance covariates.

For scanned Player 1 participants the statistical model included regressors corresponding to the onset of the Decision period for *Normal round coordinate decisions*, *Normal round non-coordinating decisions*, *Special round deviate decisions* (where the participant chose to deviate from the established coordinating response); and *Special round coordinate decisions* where the participants chose to maintain the established convention in spite of the incentive to deviate. For Player 2 participants (who were not shown an altered payoff matrix on Special rounds) regressors were generated for *Coordinating decisions* and *Non-coordinating decisions*.

Separate regressors corresponding to the Outcome period for both Players 1 and 2 participants included *Normal round coordinate outcomes*, *Normal round non-coordinate*

fMRI coordination game study

outcomes (post convention being established); Normal round non-coordinate outcomes (pre coordinating convention); Special round coordinate outcomes and Special round deviate outcomes.

One sample t-tests on the statistical regressors of interest were generated for each participant and entered in the 2nd level (“random effects”) analysis to look for statistically significant patterns of activation across participants as a group. Decision period activity for Player 1 participants was analysed using a 2 x 2 factorial model with round type (Normal / Special) and decision (coordinate / deviate) as factors. A 2 x 2 factorial model was applied to outcome period activations for the Player 2 participants for coordinate and non-coordinate Special and Normal round outcomes. Coordinate and non-coordinate Normal round outcomes were also compared using a two sample t-test which combined data from both Player 1 and Player 2 participants. Covariates were also included in the 2nd level models to control for inter-subject variance arising from two factors: the number of coordinating rounds achieved prior to the first Special round and whether or not the participant did or did not meet the other player in the waiting room before hand (see Participants above and Behavioural results below).

Six participants were excluded from the 2nd level fMRI factorial analysis described below either to bad fMRI data (2 participants) or a failure to make at least one example of the modelled event types on Special or Normal rounds (e.g. if all Special rounds resulted in coordinating or non-coordinating outcomes). The 2nd level factorial analysis was therefore based on results from 15 Player 1 participants and 11 Player 2 participants. Results are reported thresholded at $p < 0.001$ uncorrected for multiple comparisons. Activation coordinates are transformed from MNI into Talairach space (Talairach and Tournoux 1988) according to the “mni2tal” transformation (www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html). Brodmann area (BA) reference numbers and anatomical labels are also given for each reported activation focus.

RESULTS

Behavioural results.

Choice behaviour

fMRI coordination game study

All participant pairs achieved a behavioural equilibrium i.e. a coordinating convention prior to the first Special round. The number of **successful** coordinating Normal rounds prior to the first Special round ranged between 2 and 8 with a mean of 6.5 across all participant pairs. For the majority of pairs both coordinate and deviate decisions occurred on Special rounds within a single run. Across all participant pairs Player 1 opted to deviate from the established convention on 45.3% of Special rounds. A trend was present towards increased rates of deviation on later Special rounds, but this was not statistically significant (46.9, 40.7, 46.9 and 53.1% for the 1st to 4th Special rounds respectively $F(3,90)=0.303$, $p=0.80$)(Figure 2a). Whether or not Player 1 participants met the other player face to face in the waiting room prior to the experiment had a significant effect on the number of deviations the Player 1 participant made on Special rounds. Compared to those who were playing against an unseen confederate (see Participants section), participants who had met the other player were significantly less likely to deviate on special rounds (mean of 2.89 deviations compared to 1.43 deviations; $t=3.35$, $p=0.002$). The factor of whether or not participants met the other player was included as a between-subjects covariate in the 2nd level random effects analysis of the fMRI data to control for differences in activity due solely to this factor (see fMRI Methods and Analysis Procedure).

As well as the expected periods of non-coordination on early Normal rounds we also observed a minority of non-coordinating outcomes on later Normal rounds which breached the established coordinating convention (8% of all Normal rounds). In 11 out of the 32 participant pairs (**4% of Special rounds overall**) a deviation by Player 1 still resulted in a **“coordinating” outcome due to a spontaneous breach in the coordinating convention by Player 2**. In these cases Player 1’s decision was still classed as a deviate decision for the sake of the fMRI analysis. The occurrence of non-coordinating Normal outcomes made by both players **after a coordinating convention had been established** may constitute an example of the so-called “trembling hand” effect in economic games (Selten, 1988) whereby participants occasionally deviate from an established optimum equilibrium without apparent explanation.

Previous behavioural studies have shown the number of prior coordinating rounds to have an effect on the likelihood of participants deviating on Special rounds in a 3 player version of the coordination game (Guala and Mittone, 2010). However, in the current study the number of prior coordinating rounds did not significantly affect the likelihood of participants deviating on

fMRI coordination game study

Special rounds (although see below). However, the number of pre-Special coordinating rounds was also included as a between subjects covariate in the main 2nd level random effects analysis of fMRI data to control for this factor (see fMRI Methods and Analysis Procedure).

Finally only 2 of the Player 2 participants chose to deduct money from Player 1 at the end of the session. Due to the small number of such deduction decisions **and the fact that participants were only faced with one such decision at the end of the session**, the punishment phase was not modeled in the fMRI statistical analysis.

Decision Times

We also analysed the time taken for participants to make decisions on the different round types and for the different decision types. Decision times were not found to differ significantly for Special rounds on which Player 1 choose to deviate relative to coordinate ($F(1,12)=0.27$). However, decision times were significantly increased for the first relative to subsequent Special round decisions in Player 1 ($F(2,24)=23.30, p<0.0001$). Decision times on special rounds were also found to be significantly affected by the number of preceding Normal rounds coordinations achieved by the participant pair, with Players that successfully coordinated on all of the preceding Normal rounds showing longer Decision Times relative to those who achieved <8 Coordinating Normal rounds prior to the first special round ($F(1,16)=16.61, p=0.027$) (Figure 2b).

fMRI results

Decision period activity

Special / Normal rounds (Player 1): A two sample t-test examining voxels which showed significantly more activity for Special round decisions relative to Normal round decisions revealed large activation clusters (>20 voxels) in the mid-brain (2,-24,-23), anterior cingulate gyrus BA32 (-14,39,-4) and BA24 (-4,29,-1); The head of caudate nucleus (-12,19,-6); parahippocampal gyrus BA37 (30, -45,-10), Inferior frontal gyrus / anterior insula (BA47) (24,7,-17); Medial prefrontal cortex BA10 (32,52,-9) and right dorsolateral prefrontal cortex / middle frontal gyrus (BA9) (34,38,29).

Coordination / deviate decisions (Special rounds Player 1): We also compared activity on Special rounds in the period leading up to either coordinate or deviate decisions. Table 1a and Figure 3a illustrates the results for Special rounds where Player 1 opted to continue coordinating, illustrating activation in the anterior insula, amygdala and orbitofrontal cortex in advance of these decisions. The opposite contrast, examining regions which showed increased activity for deviate relative to coordinate decisions included activity in the bilateral head of caudate nucleus and orbitofrontal cortex as well as a strong locus of activity in the mid-brain (Table 1b Figure 3a).

Outcome period activity

Special rounds (Player 2): Outcome period activity in Player 2 for rounds on which Player 1 opted to coordinate or deviate were examined using a two sample t-test. Comparison of coordinate versus deviate outcomes revealed a large cluster of activation in the rostro-medial prefrontal cortex corresponding to bilateral BA10 (-3,59,-8; 25,51,17) and BA9 (-23,28,24; 13,49,19). The reverse contrast comparing deviate and coordinate round outcomes highlighted a different locus within the rostro-medial prefrontal cortex (-15,36,-3) along with the anterior cingulate gyrus BA32 (21,44,5). Despite further examination of these contrasts at a reduced statistical threshold ($p < 0.01$ uncorrected), Special round outcome activity following Player 1 deviations was not observed in regions previously linked to negative affective processing as predicted by our apriori hypotheses (e.g. amygdala / anterior insula, see introduction).

Normal round outcomes (Players 1 and 2): Outcome period activity was also compared on Normal rounds in the cases where the other player either spontaneously deviated from an established norm or continued coordinating with the stable convention (see Behavioural results). Regions with increased activity for coordinate relative to deviate outcomes included the amygdala, medio-dorsal thalamus and parahippocampal gyrus as well as the mid-brain (Table 2a, Figure 3b). For the deviate relative to coordinate Normal round outcome comparison significant activity was observed in the bilateral anterior insula, along with the midbrain (substantia nigra) and **rostral medial** prefrontal cortex (Table 2b; Figure 3b).

DISCUSSION

This is the first study to examine brain areas involved in interactive decision making in a social dilemma task embedded in the context of a longer series of coordination tasks. Previous studies that ignored the “historical” context of social behaviour had focused on decisions concerning whether to share a sum of money fairly (Sanfey et al., 2003; Volz et al., 2009), whether to cooperate or defect (Rilling et al., 2002) or shirk / work in a job (Hampton et al., 2008). Based on this body of work, successful cooperation was expected to be accompanied by activity in limbic, emotional and reward centres in the brain, reflecting the positive affective value of pro-social decisions over and above their monetary reward. We also predicted that breaches in the coordinating equilibrium would be associated with activity in regions linked to negative affect such as the anterior insula and amygdala. These expectations were only partially borne out by the present results, suggesting that decision making in repeated coordination/deviation tasks may rely on different mechanisms.

As expected, the decision to take advantage of the additional payment on Special rounds was linked to mid-brain, head of caudate and orbitofrontal activation, consistent with a response to an unanticipated reward (Schultz, 2000) (Table 1b). Prior to coordinating decisions on Special rounds, some activation was present in structures implicated in reward processing (e.g. caudate nucleus), but significant activity was also observed in the amygdala and anterior insula, two areas more commonly associated with negative affective states (Table 1a; Figure 3a). One possibility is that activity in these regions reflects either an aversive response to deviating from a norm or the fear of losing an established equilibrium in the game. This may interact with the rewarding aspects of social cooperative behaviour to influence participants’ decision making on Special rounds (Rilling et al. 2002).

It should be noted however that amygdala activity was also present for the socially and financially rewarding coordinating Normal round outcomes (Table2a; Figure 3b). One possibility is that the amygdala’s role in social processing is broader than commonly assumed. As well as responding to negative stimuli it may also respond to other events which do not have a clear affective valence but are salient for social decisions (Adolphs et al., 2004). **Rather than a privileged role in processing fearful stimuli a more accurate view of amygdala function is that it plays a general role in modulating attention to and interpreting salient environmental**

events and stimuli (Small et al., 2003; Pessoa 2010), tracking positive as well as negative reward expectancies (Belova et al., 2007). This issue could be further investigated using a psycho-pharmacological approach. It has been suggested that the neuro-peptide oxytocin enhances pro social trust (Fehr et al. 2008) by down regulating activity in the amygdala (Baumgartner et al. 2008). However, if amygdala activity on Special rounds truly reflects fear of the consequences of breaking a social norm, then oxytocin administration should attenuate this response leading to an *increased* rate of deviation in the current game.

A supplementary goal of the present experiment was to investigate the effect of previous coordination experience. Previous work using a 3 player version of a simple coordination game has provided evidence that the number of preceding coordinating rounds can affect the likelihood of participants deviating on Special rounds (Guala and Mittone, 2010). This is consistent with the idea that arbitrary coordinating conventions “intrinsic” to the task can develop into social norms via repeated execution (see introduction). We did not find a significant effect of prior coordination on the likelihood of Players coordinating or deviating in the current study. However, analysis of the time taken for participants to make decisions showed that an extended period of coordination prior to first Special round affected decision time, such that decision times were longer for participant pairs who had coordinated on all 8 of the preceding Normal rounds relative to those that did not.

The number of coordinating rounds prior to the 1st Special round was modelled as a “nuisance” covariate in the main factorial analysis of fMRI activity (see fMRI methods above). However, given the behavioural effects described above and reasoning from social and evolutionary philosophy (see introduction), it might be expected that activation of brain areas associated with pro-social motives and cognitive functions should be stronger when social “groupness” and social norms have been reinforced by repetition. We also therefore carried out a supplementary analysis in which the effect of this factor on activity during the decision period on Special rounds was assessed. Activity in the bilateral amygdala increased significantly with the number of preceding coordinations, consistent with this region activating in response to anticipated breaches in social norms (Figure 4a). In contrast, activity in the rostral medial prefrontal cortex was increased for participant pairs who achieved fewer coordinations prior to the first Special round, consistent with enhanced demands on mentalising processes in these participants (Figure 4b). This result suggests that repeated

execution of an arbitrary convention may modify the brain response to Special round decisions later in the game in a manner consistent with our supplementary hypotheses.

Anterior insula activity had been predicted for outcome periods when Player 2 saw that Player 1 had deviated on Special rounds, consistent with activation of a “social disgust” response (Sanfey et al. 2003). However this effect was not found. Instead, both coordinate and deviate outcomes mainly activated rostral medial prefrontal cortical areas (Table 2a). A key difference in the current relative to previous studies is that the outcome screen revealed Player 1’s altered payoff structure to Player 2. The reasons for Player 1’s deviations (i.e. the additional monetary incentive) were therefore apparent to Player 2 following such decisions. It is possible that because of this, the strong affective response in participants was disrupted by the inception of higher cognitive processes devoted to rationalizing the surprising result (Hariri et al., 2000; Eisenberger and Lieberman, 2004). The strong locus of activity within rostral medial prefrontal regions during these outcome periods is consistent with this explanation as activity in these regions has previously been implicated in mentalising processes. Further, the great majority of Player 2 participants opted *not* to exact a financial punishment on Player 1 at the end of the task, suggesting that they did not deem deviations on Special rounds as a breach of deep seated social norms.

Interestingly, when deviations *did not* have a clear rationale to the observer (i.e the payoff matrix shown during the outcome period revealed that there was no monetary incentive for the deviation), activity was observed in the bilateral anterior insula (Figure 2b). Spontaneous deviations occurred on Normal rounds in most participant pairs even when a coordinating equilibrium had been established (see Behavioural Data). The reasons for players making these decisions are not clear, although the phenomenon of players spontaneously deviating from game equilibria has been explored elsewhere (Selten, 1988). What is certain is that the outcomes on these rounds were not expected by the other player who had opted to continue with the mutually rewarding strategy established on the preceding rounds. These inexplicable “norm” breaches were also associated with activity in the mid-brain (substantia nigra) and rostral **medial** prefrontal cortex. Together these structures may be important in generating a social “prediction error” which mediates associative learning in social contexts (Table2b)(Hampton et al. 2008).

fMRI coordination game study

Although the above discussion has emphasised emotional factors as contributing to socially normative behaviour it is also likely that purely cognitive mechanisms contribute to adherence to social norms. For example, the observation of an action in another individual may itself provide a strong cue to initiating the same behaviour in the observer without the need to evoke the concept of a social norm (Gallese et al., 2004; Ochner et al., 2008; Bastiaansen et al. 2009). The anterior insula activity reported for Player 1 decisions to coordinate on Special rounds was located in the most superior and anterior part of this structure (BA47) which is contiguous with the inferior frontal gyrus, a region hypothesised to contain “mirror neurons” in humans (Rizzolatti and Craighero, 2004). Further studies could test the idea that norm adherence is a sophisticated form of response “mirroring” using a payoff structure which incentivised participants to adopt a “non-mirroring” convention (i.e. both players rewarded for a different response). This would eliminate the influence of imitative behavioural tendencies on decision making in the task. It would also eliminate the possible influence of the existing social norm of conformity on Special rounds. The role of this task-extrinsic norm is potentially problematic for the idea that the coordination game is a better model of the evolution of social norms than for example the prisoner’s dilemma game. However, by modifying the payoff structure such that Player 1 would be incentivised to deviate by making the *same* response as the other participant the role of a social conformity bias could be directly assessed. Another process which has previously been linked to norm adherence is response inhibition. Consistent with this mechanism being involved in norm adherence, both the dorsal and ventro-lateral prefrontal regions were more active in advance of Special rounds where Player 1 chose not to take an increased payout and continued coordinating (Table 1a) (Aron, 2004; Sanfey et al., 2003; Knoch et al., 2006; Spitzer et al., 2007).

Recently a general framework has been proposed for understanding the neuroscience of value-based decision making (Rangel, Camerer, Montague, 2008). According to this account at least three different levels of hierarchical control can influence choice behaviour. A “Pavlovian” system controls automatic behaviour triggered by a narrow range of stimuli and largely independent from the actual needs of the organism (for example, grabbing food even if not hungry). An “Habitual” system in contrast learns associations between positive/negative outcomes and a larger repertoire of behaviour (for example receiving a reward for coordinating with your partner). These learned behavioural drives can come into conflict with “Goal-directed” control which seeks to realise rewards by computing the value of choices in a new

fMRI coordination game study

situation, for example a change in the outcome contingencies in a strategic game. These three systems are suggested to work together to determine the values of competing behavioural choice options confronted by humans, but an open question is how social and moral norms influence valuation of choice options within such a framework.

One idea would be that cooperative norms arise from repeated stimulus-response associations learned by the habit based system, which over time acquire additional value within the goal-directed system over and above their immediate material and monetary value. The observed mid-brain activity during decision and outcome periods on Special rounds for which Player 1 deviated is consistent with the hypothesis that outcome-value computations are evoked during conformity to social norms. The modulating effect of the number of prior coordinations on Normal rounds in activity during Special round decisions (see above and Figure 4) could also be interpreted as reflecting the strength of a learned coordinating “habit”. Nevertheless, although such speculation is of theoretical interest we acknowledge there is no direct evidence in our data that can cleanly separate the role of habitual and goal-directed choice systems or establish that coordinating conventions constitute a learned habit.

Indeed, we suggest that extreme caution should always be exercised in making “reverse inferences” linking fMRI activations to cognitive and emotional processes. The role of the OFC and striatum in value based decision making is still controversial as are the roles of the amygdala and rostral medial prefrontal cortex in social and emotional processing. In the case of the OFC for example, some studies have implied that the lateral / medial division in the structure relates to outcome valence (ODoherty et al. 2001; Small et al. 2003), with lateral areas coding negative outcomes and medial areas responding to positive outcomes. However, other work has shown that the medial OFC maintains steady stimulus–outcome associations, whereas the lateral OFC represents changing (unsteady) outcomes irrespective of valence, such that it is particularly unexpected outcomes cause lateral OFC activity (Windmann et al., 2006).

One potential criticism of the current design is that on Special rounds participants received additional informational content in the decision period compared to the more frequent Normal rounds. Activations observed for the contrast of Special versus Normal rounds may at least partially reflect the increased information processing demands of Special rounds relative to

Normal rounds rather than the social / affective decision making demands of the task. However, such an account would not be able to explain the observed differences in activations dependent upon Player 1's decision to coordinate or deviate on these rounds (Table 1; Figure 3a) or the observed correlation between activity and the preceding number of successful coordinations (Figure 4). Similarly for outcome related activity the differences in activity observed decisions to either deviate or coordinate on Special rounds cannot be explained simply by the increased amount of text to read or information to process as these were equivalent for the two types of trial outcome (Figure 3b). In all these cases the information demands of the trials being contrasted were therefore matched whilst the decision making processes or observed outcomes varied.

Another potential methodological weakness of the current procedure was the low number of Special rounds used in each participant pair (4 trials). This was done as we were concerned that repeated presentation of Special rounds with altered payoffs would dilute the normative status of the response convention established by players in the preceding rounds. Although it is more common practice to average fMRI data over many more trials in each subject, there is no commonly accepted minimum number of trials per condition for effective fMRI investigations, and some studies have successfully measured the BOLD response to single trials in single subjects (Posse et al. 2001; Michelli, Price, Henson, Friston, 2003). The minimum number of trials required to measure a given BOLD response depends upon the size of the effect to be measured, together with the number of participants tested and the inherent variability of the measured signal (all of which factor together to determine the statistical power of a given study). As we could not be sure of the expected size of activations under each condition a priori, the low trial numbers meant that a risk for the current study was Type II statistical errors whereby we failed to detect activations which were in fact present in our participants. In the light of this we report activations uncorrected for multiple comparisons (see Lieberman, & Cunningham, 2009 for a justification). Significant multi-voxel clusters of activity were observed in all the main contrasts between trial types and outcomes of interest, suggesting that the study did not suffer from excessive Type II statistical errors. Nevertheless, given that the rates of deviation did not differ significantly from the 1st to the 4th round, we conclude that future fMRI studies may safely increase the number of Special rounds to increase statistical power without compromising the nature of the coordination game.

fMRI coordination game study

Finally, it is interesting to directly compare the results of the current fMRI study with those reported for the Prisoners Dilemma (PD) game, which is a more commonly used model task for studying adherence and deviation from social cooperative norms (Rilling et al. 2002). In the PD game, activity in the “antero-lateral striatum” and ventro-medial / Orbitofrontal cortex was found to be increased during decisions to “Cooperate”, presumably reflecting the anticipated social and monetary reward of the outcome. Both the OFC and antero-lateral region were maximally activated by Cooperate-Cooperate trial outcomes in the PD. Interestingly, our results show striatal activity very close to the region described as the antero-lateral striatum in Rilling et al.’s study (labelled “Head of Caudate” in Tables 1 and 2 below). Activations were also observed in this region in advance of Player 1 decisions to deviate (i.e. obtain additional monetary reward) as well as during observation of non-coordinating outcomes on Normal rounds. In contrast to the PD game however OFC activation in the present study was located in the lateral rather than medial part of the structure (see discussion of lateral versus OFC activity above).

SUMMARY AND CONCLUSIONS

In common with other studies which have used tasks in which participants decide whether to adhere or deviate from social norms, we find evidence for activity in limbic brain structures which mediate approach and avoidance behaviours during a coordination game. Our analysis suggests that it may be the anticipated negative consequences of deviation that sustain coordination as well as the rewarding aspects of adhering to behavioural convention. The findings also indicate that norm breaches may only elicit an aversive response when they have no rational explanation for the recipient and that the extent of previous periods of coordination on a convention may affect the brain response when participants consider breaking the established “norm” later in the game. We suggest that coordination games of the type used here have considerable potential for future explorations of norm driven behaviour and brain processes in humans.

REFERENCES

- Adolphs R, Spezio M, Anders S, Ende G, Junghoffer M, Kissler J and Wildgruber D (2006) Role of the amygdala in processing visual social stimuli. Symposium on Understanding Emotions SEP, 2004; Understanding Emotions, 156, p 363-378.
- Aron, A.R., Robbins, T.W., and Poldrack, R.A. (2004). Inhibition and the right inferior frontal cortex. Trends Cogn. Sci. 8, 170–177.
- Barbey AK, Krueger F, Grafman J (2009) An evolutionarily adaptive architecture for social reasoning. Trends In Neurosciences, 32(12), p 603-610.
- Bastiaansen JACJ, Thioux M, Keysers C (2009) Evidence for mirror systems in emotions. Philosophical Transactions Of The Royal Society B-Biological Sciences, 364(1528), 2391-2404.
- Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U and Fehr E (2008) Oxytocin shapes the neural circuitry of trust and trust adaptation in humans: Neuron, 58(4), p 639-650.
- Belova, MA, Paton, JJ, Morrison, SE, & Salzman, CD (2007). Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. Neuron, 55, 970–984.
- Bicchieri C (2006) The Grammar of Society: The Nature and Dynamics of Social Norms (Cambridge University Press, 2006).
- Binmore K (1998) Game Theory and the Social Contract: Volume 2: *Just Playing*. Cambridge: MIT Press.
- Binmore K (2005). Natural Justice. New York: Oxford University Press.

fMRI coordination game study

De Quervain DJF, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A and Fehr E (2004) The neural basis of altruistic punishment. *Science*, 305, 1254-1258.

Eisenberger NI & Lieberman MD (2004) Why rejections hurt: a common neural alarm system for physical and social pain. *TRENDS in Cognitive Science* 8, 294-300.

Elster J (1998) Emotions and economic theory. *Journal Of Economic Literature*, 36(1), p47-74.

Fehr E, Pfaff D, Kordon C, Chanson P and Christen Y (2008) The effect of neuropeptides on human trust and altruism: A neuroeconomic perspective *Conference on Hormones and Social Behavior Dec 2007; Hormones and Social Behaviour*, 47-56

Fischbacher U (2007) Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171-178.

Gallese V, Keysers C and Rizzolatti G (2004) A unifying view of the basis of social cognition. *Trends In Cognitive Sciences*, 8(9), 396-403.

Gintis H (2009). *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences* (Princeton University Press).

Greene JD, Nystrom LE, Engell AD, Darley JM and Cohen JD (2004) The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), p 389-400.

Guala F and Mittone L (2010) How History and Convention Create Norms: An Experimental Study. *Journal of Economic Psychology*, 31, 749-756.

Hampton AN, Adolphs R, Tyszka MJ and O'Doherty JP (2007) Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. *Neuron*, 55, 545-555.

Hampton AN, Bossaerts P and O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences (USA)*, 105(18), 6741–6746.

fMRI coordination game study

Hare TA, Camerer CF, Rangel A (2009) Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System. *Science*, 324(5927), 646-648.

Hariri AR, Bookheimer SY, Mazziotta JC (2000) Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport* 11, 43-48.

Haruno M & Frith CD (2010) Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, 13(2), 160-161.

Henrich J & Henrich N (2007) *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. New York: Oxford University Press.

Hume D (1748) *An Enquiry concerning Human Understanding*.

Klucharev V, Hytonen K, Rijpkema M, Smidts A, Fernandez G (2009) Reinforcement Learning Signal Predicts Social Conformity. *Neuron*, 61(1), 140-151.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832.

Krueger, F; Barbey, AK; Grafman, J (2009) The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, 13(3), 103-109.

Lieberman, M. D. & Cunningham, W. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, 4, 423-428.

Lewis DK (1969) *Convention: A Philosophical Study*. Cambridge, Mass.: Harvard University Press.

Mehta, J., Starmer, C. and Sugden, R. 1994. The Nature of Salience: An Experimental Investigation of Pure Coordination Games. *American Economic Review*, 84: 658-673.

fMRI coordination game study

Michelli A, Price CJ, Henson RNA, Friston KJ (2003) Estimating efficiency a priori: a comparison of blocked and randomized designs. *Neuroimage*, 18, 798-805.

Montague PR and Lohrenz T (2007) To detect and correct: norm violations and their enforcement. *Neuron*, 54, 14-18.

O'Doherty J, Kringelbach ML, Rolls ET, Hornak J and Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, 4, 95-102.

O'Doherty, JP, Deichmann R, Critchley HD and Dolan RJ (2002) Neural responses during anticipation of a primary taste reward. *Neuron*, 33, 815–826.

Ochsner KN, Zaki J, Hanelin J, Ludlow DH, Knierim K, Ramachandran T, Glover GH, Mackey SC (2008) Your pain or mine? Common and distinct neural systems supporting the perception of pain in self and other. *Social Cognitive And Affective Neuroscience*, 3(2), 144-160.

Pessoa L (2010) "Emotion and cognition and the amygdala: From "what is it?" to "what's to be done?" *Neuropsychologia* 48(12), , pp. 3416–3429.

Posse S, Binkofski F, Schneider F, Gembris D, Frings W, Habel I, Salloum JB, Mathiak K, Wise S, Kiselev V, Graf T, Elghahwagi B, Grosse-Ruyken ML and Eickermann T (2001) A new approach to measure single-even related brain activity using real-time fMRI: Feasibility of sensory, motor and higher cognitive tasks. *Human Brain Mapping* 12, 25-41.

Phillips ML, Young AW, Senior C, Brammer M, Andrew C, Calder AJ, Bullmore ET, Perrett DI, Rowland D, Williams SCR, Gray JA, David AS (1997) A specific neural substrate for perceiving facial expressions of disgust. *Nature*, 389, 495-498.

Rangel A, Camerer C and Montague PR (2007) A Framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*. 9. 545-556.

fMRI coordination game study

Rilling JK, Gutman DA, Zeh TR, Pagnoni G, Berns GS, & Kilts CD (2002) A Neural Basis for Social Cooperation. *Neuron*, 35, 395–405.

Rizzolatti G, Craighero L (2004) The mirror-neuron system. *Annual Review Of Neuroscience*, 27, 169-192.

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE and Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. *Science*, 300, 1755-1758.

Schelling (1960) *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Schultz W (2000) Multiple reward systems in the brain. *Nature Reviews Neuroscience*, 1, 199–207.

Selten, R. (1988) Evolutionary stability in extensive two-person games - correction and further development. *Mathematical Social Sciences*, 16:223–266.

Skyrms B (1996) *Evolution of the Social Contract*. Cambridge University Press.

Skyrms B (2004) *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press.

Small D M, Gregory MD, Mak YE, D Gitelman D, Mesulam MM & Parrish T (2003) Dissociation of Neural Representation of Intensity and Affective Valuation in Human Gustation. *Neuron*, Vol. 39, 701–711.

Spitzer M, Fischbacher U, Herrnberger B, Gron G and Fehr E (2007) The neural signature of norm compliance. *Neuron*, 56, 185-196.

Sugden R (1986) *The Economics of Rights, Cooperation and Welfare*. Oxford: Blackwell.

Sugden R (1989) Spontaneous Order. *Journal of Economic Perspectives*, 3, 85-97.

fMRI coordination game study

Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotactic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging* (Stuttgart: Thieme).

Ullmann-Margalit E (1977) *The Emergence of Norms*. Oxford: Clarendon Press.

Volz, K.G., Kessler, T., & von Cramon, D.Y. (2009). In-group as part of the self: In-group favoritism is mediated by medial prefrontal cortex activation. *Social Neuroscience*, 4, 244-260.

Windmann S, Kirsch P, Mier D, Stark R, Walter B, Güntürkün O, and Vaitl D (2006) On Framing Effects in Decision Making: Linking Lateral versus Medial Orbitofrontal Cortex Activation to Choice Outcome Processing. *Journal of Cognitive Neuroscience* 18:7, pp. 1198–1211

fMRI coordination game study

a)

Cluster size	Anatomical area	T-statistic	x	y	z
78	Body of Caudate	5.90	-12	11	22
254	Anterior insula / Inferior frontal gyrus BA47	5.56	-20	29	-10
	Middle frontal gyrus BA11	4.07	-22	36	-10
160	Pulvinar nucleus	5.40	-4	-35	5
	Pulvinar	4.62	6	-33	7
118	Inferior frontal gyrus BA13	5.36	38	9	-12
	Superior Temporal gyrus BA38	3.94	32	7	-21
52	Body of Caudate	5.10	10	9	16
87	Parahippocampal gyrus BA36	4.97	-38	-34	-10
	Parahippocampal gyrus	3.49	-38	-22	-11
100	Hypothalamus	4.38	-2	1	-12
	Globus Pallidus	4.37	-14	0	-8
17	Medial frontal gyrus BA6	4.36	-6	-22	64
28	Tail of Caudate	4.25	26	-42	15
41	Parahippocampal gyrus BA28	4.22	26	-18	-8
	Hippocampus	3.69	28	-20	-16
	Amygdala	3.59	22	-12	-11
19	Superior frontal gyrus BA9	4.00	-18	54	25
13	Middle frontal gyrus BA11	3.91	-36	56	-10
18	Cerebellum anterior lobe	3.65	2	-44	-25

b)

Cluster size	Anatomical area	T-statistic	x	y	z
300	Midbrain - Red Nucleus	7.09	2	-24	-4
804	Medial Frontal Gyrus BA8	6.11	14	35	42
	Anterior Cingulate BA24	5.83	12	19	21
328	Posterior cingulate Gyrus BA31	6.01	18	-35	33
184	Postcentral Gyrus BA2	5.82	-38	-26	29
98	Parietal Lobule BA40	5.60	-59	-37	46
88	Anterior Cingulate BA25	5.59	2	9	-6
	Right Caudate Head	4.02	2	0	2
103	Cingulate Gyrus BA24	5.23	6	-3	24
	Cingulate Gyrus BA23	4.13	-2	-10	26
38	Middle Temporal Gyrus BA21	5.13	65	-12	-11
25	Globus Pallidus	4.97	-16	-1	9
	Thalamus Vento-Lateral Nucleus	3.48	-14	-9	8
95	Orbitofrontal cortex BA10	4.94	32	52	-11
	Superior Frontal Gyrus	3.91	20	54	-9
36	Cerebellum Posterior Lobe	4.86	-32	-68	-30
49	Inferior Frontal Gyrus	4.81	24	7	-17
75	Cingulate Gyrus BA31	4.81	-8	-48	41
108	Anterior Cingulate BA 24	4.77	-10	27	-3
	Left Caudate Head	4.17	-12	19	-6
76	Medial Frontal Gyrus BA10	4.76	10	42	-12
	Anterior Cingulate BA32	3.71	18	38	-7
73	Parahippocampal Gyrus BA37	4.74	30	-45	-10
	Fusiform Gyrus	4.03	34	-36	-15
48	Cerebellum (Culmen)	4.62	6	-38	-20
26	Middle Temporal Gyrus	4.33	61	-28	-10
27	Lentiform Nucleus Putamen	3.78	20	19	-3
61	Middle Temporal Gyrus BA22	4.06	59	-33	3
	Superior Temporal Gyrus BA41	3.57	55	-29	9

Table 1 Talairach coordinates of major activations clusters in the Decision period on Special rounds for Player 1. a) Activations during sustained coordination compared to deviation decisions. b) Deviation compared to coordination decisions on Special rounds.

fMRI coordination game study

a)

Cluster size	Anatomical area	T-statistic	x	y	z
88	Amygdala	5.29	23	-8	-10
	Hippocampus	4.09	25	-16	-18
44	Midbrain	5.01	-3	-33	0
34	Medio-dorsal thalamus	4.42	-2	-17	6
25	Anterior nucleus Thalamus	4.29	-9	14	14
29	Posterior cingulate BA31	4.27	-2	-25	40
14	Parahippocampal gyrus BA30	3.93	23	-37	4

b)

Cluster size	Anatomical area	T-statistic	x	y	z
471	Orbitofrontal cortex BA11	6.65	-28	44	-11
1350	Fusiform Gyrus BA20	6.56	-40	-13	-25
	Middle Temporal Gyrus BA 21	5.54	-57	3	-14
1120	Midbrain	6.39	-2	-30	-19
	Midbrain (Substantia Nigra)	5.78	12	-18	-7
350	Right Cerebellum	6.36	30	-68	-32
183	Superior Temporal Gyrus BA39	5.80	-40	-48	13
	Parahippocampal Gyrus BA19	4.67	-40	-48	2
	Superior Temporal Gyrus BA39	4.04	-50	-54	6
829	Superior Temporal Gyrus BA38	5.74	46	11	-16
189	Precentral Gyrus BA6	5.59	-44	-16	60
	Precentral Gyrus BA6	4.49	-24	-10	71
162	Inferior Parietal Lobule BA40	5.52	50	-58	36
140	Middle Temporal Gyrus BA21	5.48	36	4	-32
926	Inferior Parietal Lobule BA40	5.37	48	-35	40
825	Superior Frontal Gyrus BA8	5.31	18	29	45
	Superior Frontal Gyrus BA10	4.97	8	30	48
246	Posterior Cingulate Gyrus BA23	5.27	4	-12	32
81	Inferior Frontal Gyrus BA47	5.20	55	18	1
	Superior Temporal Gyrus BA22	4.42	61	10	-2
166	Inferior Parietal Lobule BA40	5.12	42	-30	29
25	Medial Frontal Gyrus BA10	5.02	16	38	-7
34	Middle Frontal Gyrus BA8	4.84	-32	14	38
193	Supramarginal Gyrus BA40	4.83	-57	-46	19
	Superior Temporal Gyrus BA22	4.61	-63	-42	15
31	Middle Temporal Gyrus BA37	4.79	-40	-60	0
124	Right Caudate Head	4.72	8	25	1
	Left Caudate Head	4.72	-2	23	3
	Right Anterior insula BA47	3.82	14	23	-13
25	Left Anterior insula BA47	4.72	-55	27	-1
65	Middle Temporal Gyrus BA21	4.71	61	-41	-8
85	Superior Temporal Gyrus BA41	4.68	44	-40	8
44	Posterior Insula BA13	4.66	34	-3	24
61	Middle Frontal Gyrus BA9	4.62	-44	25	34
79	Anterior Cingulate Gyrus BA32	4.54	-18	4	40
182	Right Cerebellum (Culmen)	4.28	30	-36	-20
	Parahippocampal Gyrus BA36	3.85	36	-30	-22
42	Orbitofrontal cortex BA11	4.22	2	58	-11
28	Middle Frontal Gyrus BA6	4.17	32	-9	45
143	Left Cerebellum (Declive)	4.13	-14	-83	-19
37	Middle Temporal Gyrus BA21	4.09	61	-1	-15
44	Inferior Temporal Gyrus BA20	4.07	53	-3	-28
25	Medial Frontal Gyrus BA6	3.84	8	-13	49
29	Postcentral Gyrus BA5	3.83	22	-41	70

Table 2 Significant loci of activation during outcome periods on Normal rounds for a) Observation of coordination versus non-coordinating decision in the other player. b) Non-coordination versus coordination.

fMRI coordination game study

A)

		Normal rounds		Special rounds	
		Player 2		Player 2	
		L	R	L	R
Player 1	L	50p, 50p	0, 0	50p, 50p	£2, 0
	R	0, 0	50p, 50p	£2, 0	50p, 50p

Player 1 (Potential "Deviant")

B)

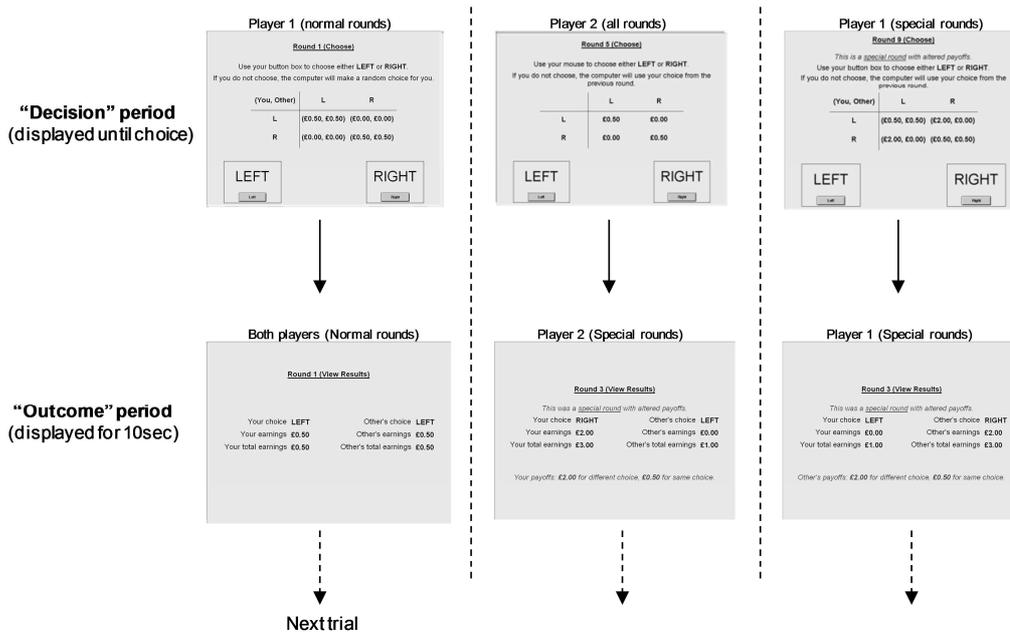
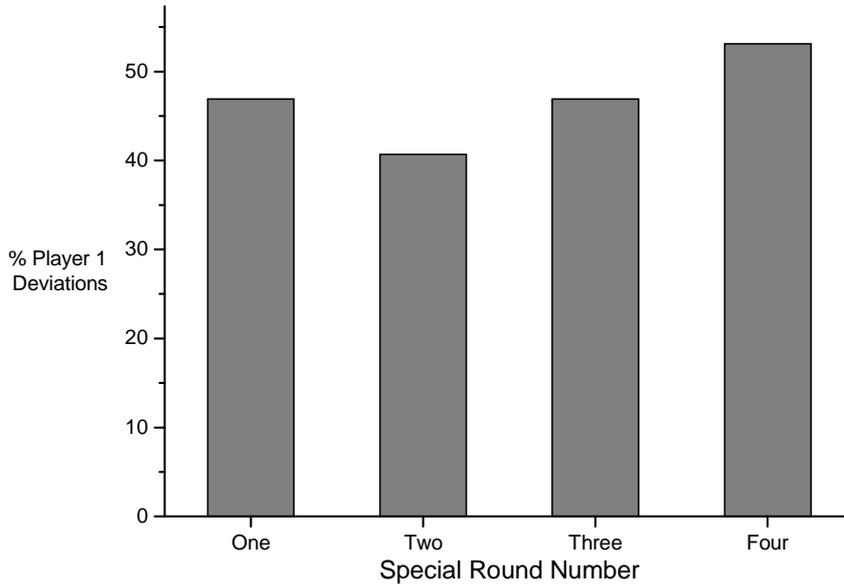


Figure 1 Task design and procedure. a) Payoff matrix for Normal and Special rounds. Player 1's payoffs are shown first. On Normal rounds there is an incentive for both players to find a behavioural convention and continue with that coordinating convention. On Special rounds there is an incentive for Player 1 to deviate from the established convention to gain additional monetary reward. b) Decision and Outcome period screens displayed to participants during the task. During the decision period both players were given information about the payoffs on that trial, with only Player 1 seeing the payoffs for both participants. During the outcome period both players were told what the actual choices and payouts were for both players, along with their total earnings for the game.

fMRI coordination game study

a)



b)

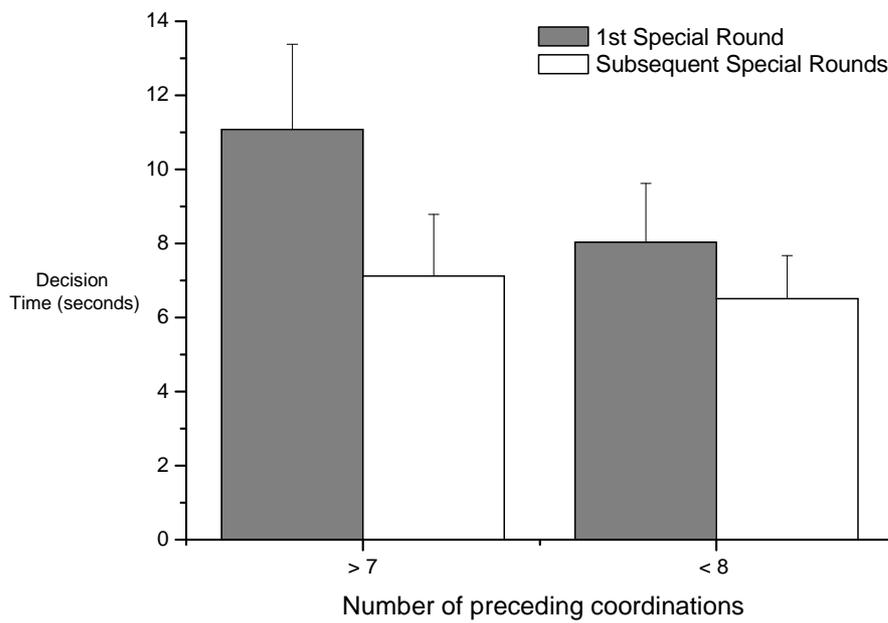


Figure 2 Behavioural data showing a) Player 1 deviation rates on the four Special rounds b) Player 1 Decision times for 1st and subsequent Special rounds plotted against number of preceding coordinating Normal round outcomes.

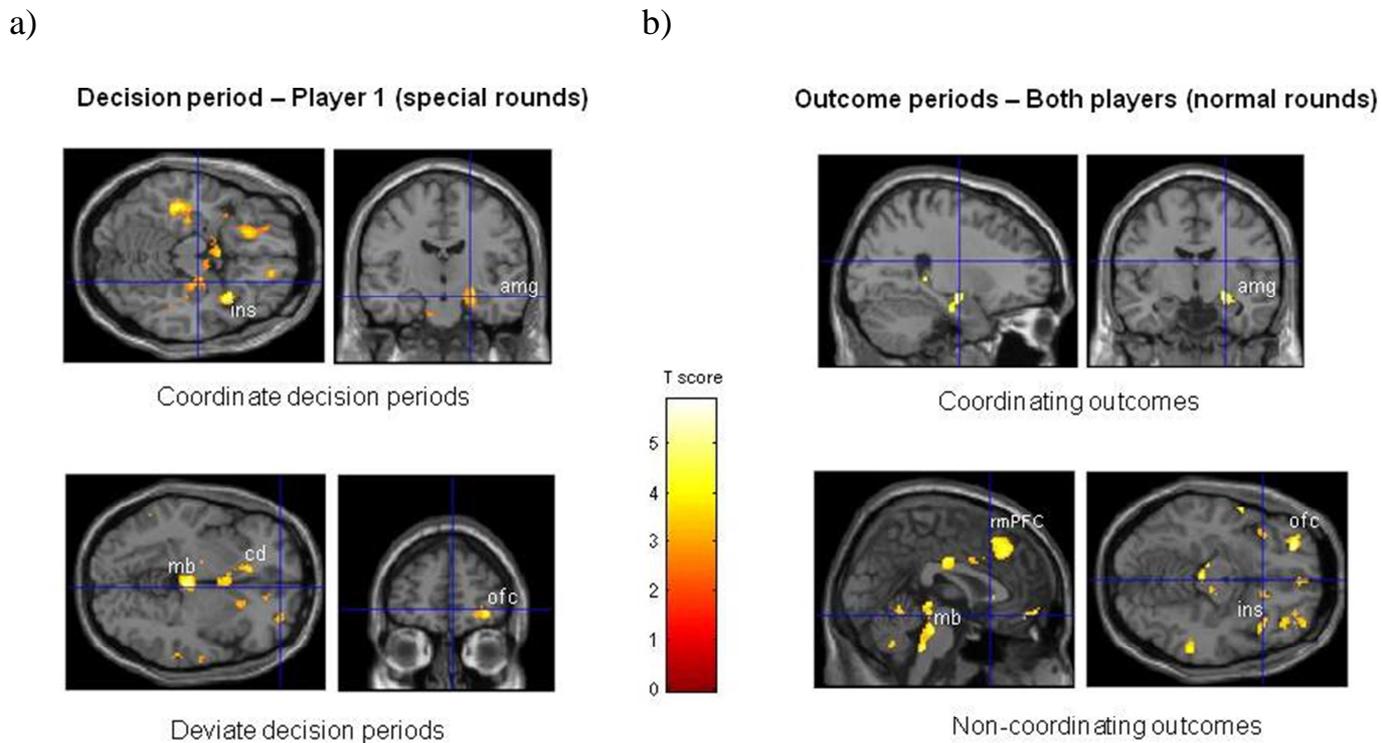


Figure 3 Activity during decision periods and outcome periods. a) During decision periods on Special rounds, enhanced activity was observed in the anterior insula (ins) and amygdala (amg) prior to coordinate decisions in Player 1 (contrast of Coordinate versus Deviate decisions for Special rounds only). In contrast when Player 1 choose to deviate from the established convention to gain additional monetary reward significant activity was seen in the mid-brain (mb), caudate (cd) and orbitofrontal cortex (ofc) compared to coordinate b) During outcome periods regions showing increased activity for Coordinating versus non-coordinating outcomes on Normal rounds included a cluster in the amygdala. Spontaneous deviations were also sometimes observed on Normal rounds even after a coordination had been established (see behavioural data). Analysis of outcome period activity in participants who observed these non-coordinating outcomes on Normal rounds showed relative increases in the bilateral anterior insula (ins), rostral medial prefrontal cortex (rmPFC), mid-brain and orbitofrontal cortex (ofc) relative to Coordinating outcomes.

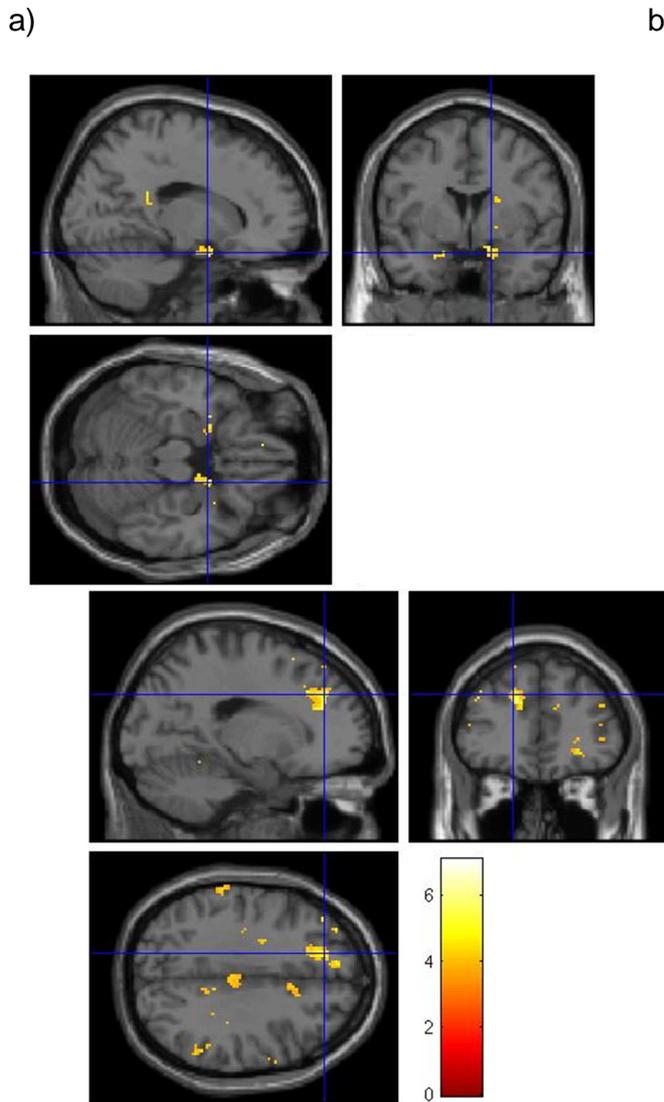


Figure 4 Results of supplementary fMRI analysis examining the change in Special round decision period activity in Player 1 with the number of Normal round coordinations achieved by each participant pair prior to the first Special Round. a) Increase in activity in bilateral amygdala with increasing number of prior coordinations. b) A region of rostral medial prefrontal cortex showing a significant increase in activity for *reduced* numbers of Normal round coordinations (t-statistic colour intensity scale as in figure 3).