Neuroeconomics of 3-Person Ultimatum Game with Voting: The Case of Responders*

Jarosław Grygolec†, Giorgio Coricelli‡, Aldo Rustichini§

February 2009

Abstract

We study responders' behavior and corresponding brain activity in 3-person ultimatum game with voting in an fMRI experiment. In the game a proposer decides the split of the pie between himself and 2 responders. If the proposed split gets majority support among 3 players it is implemented, otherwise each player gets nothing. The design has two advantages. First, it avoids priming to discriminate against one of players as in 3-person ultimatum with a dummy of Güth, van Damme (1998). Second, it allows naturally for situations to arise in which a responder receives either less or *more* than one of players. The latter is unlike in the classical 2-person ultimatum game where proposers' advantage, experimentally a sure thing, effectively prevents studying responders' social preferences in the domain of advantageous inequality of a responder. In this paper we focus on low but positive offers to a responder inside MRI machine. The analysis reveals that the decision process on such offers can be separated into two stages. We observed increased activity in nucleus accumbens (positive reward) and amygdala (negative emotion) in the early stage. We interpret this as evidence of a decision conflict, which is detected by anterior cingulate cortex and resolved with help of bilateral insula and right dorso-lateral prefrontal cortex in the late stage of offer evaluation. Based on presented evidence we suggest that the involvement of negative emotions in rejections of low but positive offers is not due to their unfairness, but rather reflects negative reputational signal that the actual acceptance can send to others about the responder.

Keywords: ultimatum game, neuroeconomics, fMRI, envy, fairness

JEL Classification: C91, D87

---

* The paper is based on the third chapter of the doctoral dissertation of Jarosław Grygolec submitted to the faculty of Graduate School of University of Minnesota, USA.

† Center for Mind/Brain Sciences - CIMeC, University of Trento, Italy. Email: jgrygolec@gmail.com.
‡ Institut des Sciences Cognitives (CNRS), Lyon, France. Center for Mind/Brain Sciences - CIMeC, University of Trento, Italy. Email: coricelli@isc.cnrs.fr.
§ Department of Economics, University of Minnesota, USA. Email: aldo.rustichini@gmail.com.

1. Introduction

In this paper we report the results of an fMRI experiment in which we scanned responders playing 3-person ultimatum game. In the game a proposer decides the split of the pie between himself and 2 responders. If the proposed split gets majority support among 3 players it is implemented, otherwise each player gets nothing. We have chosen to study 3-person ultimatum game because it allows naturally for situations to arise in which a responder receives either less or *more* than one of players. Our investigation is motivated by our previous study Grygolec, Coricelli, Rustichini (2008) in which we found neuroeconomic evidence for envy and pride effects in non-strategic environment. The key question we want to investigate is whether envy and pride effects extend to strategic situations. It seems natural to expect that envy and pride could potentially play a role in explaining subjects' behavior in ultimatum game.

The perspective differs from the interpretation of rejections of positive amounts of money observed in 2-person version of ultimatum game, most notably theories of Fehr, Schmidt (1999) and Bolton, Ockenfels (2000), which specifically emphasize the role of aversion to advantageous inequality (the opposite of pride). Recently echoing the fairness interpretation Sanfey et al (2003) studied neural basis of responder's rejections in 2-person ultimatum game. They suggest negative emotions, as measured by activity in anterior bilateral insula, to be behind rejections of positive amounts of money. They provide interpretation that negative emotional response is evoked by unfairness of offers being rejected. However, all offers in the experiment of Sanfey (2003) were with proposer's advantage of varying degree, so that their results necessarily apply to fairness understood only as, in the parlance of Fehr, Schmidt (1999), aversion to disadvantageous inequality (or simply *envy*). This is rather weak concept of fairness in contrast to the strong one that would include aversion to advantageous inequality as well. The study of 3-person ultimatum game with voting has an advantage of separating hypothesis of social comparisons versus alternative hypothesis of strong fairness. These hypotheses are, at best, difficult to distinguish within 2-person ultimatum game in the credible way. We performed MRI experiment with one responder playing VUG inside MRI machine, and the other two players: a proposer and a second responder outside MRI. We find evidence for the role of nucleus accumbens in evaluation of offers by responders in VUG, a region associated in neuroeconomic literature with measuring subjective value of rewards as well as with envy and pride effects in our study reported in Grygolec, Coricelli, Rustichini (2008). In this paper we address the issue how neural measure of monetary value as proxied by activity in nucleus accumbens is reconciled with negative emotional responses to low, but still positive offers. The previous work of Sanfey et (2003) suggested that anterior cingulate cortex is involved in detection of such decision conflict between anterior bilateral insula, which represents negative emotional response in accordance with other literature, and right dorso-lateral cortex which was connected with executive control elsewhere. Yet, the reward areas like nucleus accumbens or orbitofrontal cortex are missing from this account, so it seems to us as incomplete. We provide evidence, based on detailed analysis of temporal precedence of brain responses during offer evaluation, that decision conflict generated by offers with low but positive amount of money to a responder consists of two stages. In the early stage, just after offer is being displayed, the nucleus accumbens in general tracks monetary value of payoff attributed to a responder. In addition, if this offer is positive but low, activity in amygdala, associated with fear and negative emotion more generally, spikes even as compared to offers in which scanned responder receives nothing. We suggest that these conflicting signals coming from nucleus accumbens (positive but low reward measured) and from amygdala (negative emotional response) is the early neural signature of decision conflict of a responder facing a low but positive offer. This decision conflict is being detected in the later stage of offer evaluation by anterior cingulate cortex. In the face of the decision conflict negative emotions related to the offer are being reevaluated in bilateral anterior insula, another important region related to negative emotion. We find that activity in this region is correlated to rejection rates on low but positive offers. However, prediction of the rejection rates between subjects is improved once we consider activity in right nucleus accumbens in addition to bilateral anterior insula. At the same the impact on decisions of right anterior insula seems to be mediated by right dorso-lateral prefrontal cortex. This result seem to be in accordance with those of Knoch et al (2006, 2007) that show that suppression of right dorso-lateral preferential cortex leads to increased acceptance of unfair offers in 2-person ultimatum game. Finally, our findings suggest that negative emotional response is not related to unfairness per se. Surprisingly, bilateral anterior insula activation exhibits significantly lower activations in response to offers that give him *nothing* rather than *positive but low* amount of money. If bilateral anterior insula was reacting to perceived unfairness of offers we should expect the opposite. One possibility is that this region produces negative emotional response to offers with positive by low amount of money not because of their unfairness, but rather because of what signal about responder accepting such an offer other players may infer. Acceptance of such offers signals weak type if anything, the reputation that is unlikely to benefit anyone in bargaining situations. In this light we think that social comparisons or alternatively strong fairness may enter the neural calculus of rejections in VUG via nucleus accumbens. We leave studying these possibilities to the separate investigation.

The rest of paper is organized as follows. In section 2 we review neuroeconomic literature on 2-person ultimatum game. Section 3 presents experimental design. The discussion of behavioral results follows in section in 4. The main results of paper on neural basis rejection in VUG are in section 5. Conclusions follow in section 6.

2. Related Literature

The research on ultimatum game and it various versions was undertaken by scientists from many different fields, starting from economics, through anthropology, psychology and most recently also neuroeconomics. It is beyond our scope here to do a complete literature review, so we limit ourselves to the most relevant papers in neuroeconomics. Sanfey et al (2003) is the first paper in neuroeconomics that investigated responders in 2-person ultimatum game. They focus on analyzing brain processes behind responders' decisions on unfair offers (responder's share of 30% or less) versus fair ones. First of all responders often reject offers of unfair offers of 30% or less, while accepting overwhelmingly fair ones. This is very robust finding, replicated in many experiments, see Camerer (2003) for review. The puzzle is why anyone would decline unfair but positive offer. Sanfey et al (2003) suggest that unfair offers induce negative emotions in responders leading to a conflict between emotions (suggesting rejection of unfair offers) and cognition (positive amount of money should be accepted). This account sounds intuitive and the evidence on brain processes by Sanfey et al (2003) seems to provide some support for this interpretation. More precisely, three regions showed increased activation in the contrast of unfair vs. fair offers: anterior insula, anterior cingulate cortex (ACC), and dorso-lateral prefrontal cortex (DLPFC). The neuroscience literature associated anterior insula with disgust, ACC with conflict monitoring and DLPFC with executive control and goal maintenance. Specifically, Sanfey et al (2003) suggested that ACC increased activation during evaluation of unfair versus fair offers may reflect detection of the conflict between sticking to self-interest motive represented by increased activation in right DLPFC and emotional response represented by increased activity in anterior insula bilaterally. Sanfey et al (2003) provide evidence that activation in anterior insula in response to unfair offers is correlated with rejections rates at the subject level. At the trial level the insula has larger activation during trials with unfair offers that result in rejections as compared to acceptance. Furthermore, the interpretation provided by Sanfey et al (2003) would suggest positive relationship between activity of right DLPFC (related to maintaining self-interest) and acceptance rates of unfair offers. However, the authors don't find correlation of activity in the right DLPFC with responders' behavior. This missing correlation sparked further studies that aimed to illuminate the role of right DLPFC in responses to unfair offers, they include van't Wout et al (2005), Knoch et al (2006) and Knoch et al (2007). The first two of these studies applied repetitive transcranial magnetic stimulation (rTMS) while the third one transcranial direct current stimulation (tDCS) to disrupt brain activity in the right DLPFC. The idea here is to compare behavior of subjects between rTMS or tDCS conditions and control sham condition, when no stimulation is applied. All three studies provide evidence that disruption of right DLPFC leads to increased acceptance rates of unfair offers, which demonstrates the causal role of right DLFPC in responses to unfair offers. In addition, Knoch et al (2006) showed that there is no similar effect in the case for left DLPFC as no significant differences were observed in acceptance rates or response times between sham condition and a condition with rTMS applied to left DLPFC. Knoch et al (2006) examined also situation when offers were generated by computers and still proposers were to receive payments. The introduction of computer offers was intended to shed light to what extent behavior of subjects to reject unfair offers is driven by reciprocity or just by purely distributional considerations, and whether right DLPFC plays the role in their decisions. Favoring reciprocity explanation Knoch et al (2006) found that responders accepted unfair offers more often than similar offers from human proposers, which points to reciprocity playing an important role in explaining subject behavior. At the same time the difference between sham condition and conditions with rTMS applied to either right or left DLPFC (but not both) were not significant. Emanuele et al (2008) point to the role of serotonergic system in behavior of responders in ultimatum game. Notably, low levels of serotonin were demonstrated to be correlated with impulsivity and aggression. Serotonergic system was also connected to irritability and expression of anger. Emanuele et al (2008) find that responders with lower levels of serotonin are more likely to reject unfair offers (responder's share10%). In the other study Wallace et al (2007) provided evidence for heritability of responders' behavior in ultimatum game. Using the classic twin design they showed that correlation of acceptance thresholds is significantly higher in monozygotic versus dizygotic twins. They analyzed also a nested model of acceptance thresholds with three sub-models accounting for genetic factors as well as for common and differential environmental factors. The genetic effects accounted for 42% of variation explained by the nested model. Burham (2007) studies the effect of testosterone levels in men for responses to low offers in ultimatum game. Higher levels of testosterone in humans and other primates were connected to aggression and dominance seeking. Accordingly, Burnham (2007) finds that men that reject versus those that accept unfair offers in ultimatum game have significantly higher testosterone levels. Koenigs, Tranel (2007) performed lesion study on behavior of responders in ultimatum game using patients after damage of ventro-medial prefrontal cortex (VMPFC). This region commands substantial interest among neuroeconomists since patients with VMPFC lesions seem to score normally on IQ tests, but in contrast control poorly emotional response. VMPFC is often considered to be more caudal part of orbitofrontal cortex, which was implicated in coding reward value. Koenig, Tranel (2007) find that VMPFC patients reject unfair offers in ultimatum game more often than controls both normal subjects and patients with lesions in other prefrontal areas.

3. Experimental Design

The study was approved by University of Minnesota Institutional Review Board. We enrolled subjects from principles and intermediate courses in economics at University of Minnesota. We organized 22 experimental sessions in total, each session with 3 subjects participating. Subjects met briefly as they were arriving at the research facility, Center for Magnetic Resonance Research at University of Minnesota. No deception was used in this study.

The experimental task involved playing repeatedly 3-person ultimatum game with voting, which is the modified version of game introduced by Güth, van Damme (1998). The game was common knowledge. There are 3 players in VUG: a proposer and two responders. One of responders was scanned with fMRI scanner while playing VUG. In the game first a proposer decides how to split $30 among three players. The shares of players were the multiples[1] of $5. The responders receive messages about the proposed split, after which all players vote whether to accept or reject the split. The majority rule is used to determine the fate of

[1] In the separate purely behavioral experiment we allowed for the splits of $30 in the multiples of $1. However, 78.7% of proposals were made in the multiples of $5. For this reason we think that the restriction of offers to be multiples of $5 didn't influence the experimental results reported here.

the split. If at least 2 players (including a proposers) vote to accept then the proposed split is carried out, otherwise each player gets 0. We denote by (x, y, z) the split decided by a proposer, with x being proposer's share, while y and z are responders 1 and 2, respectively. We varied information that responders received before voting about the splits of the pie between three players, see Table 1. In treatment Full Info responders learned before casting their votes the shares of each player, or (x, y, z). In treatment Power they learned only the proposed share of the proposer x, so that they didn't learn exactly their payoffs before voting. In treatment Me responders learned their own proposed payoffs, y in case of Responder 1 and z in case of Responder 2. In the remaining treatment Other each responder learned only what the *other* responder was to get according to a split by proposer. In addition to varying information that responders obtained before voting about the proposed split, we also allowed for possibility that sometimes a computer instead of a human proposer generates the offer. The computer offers used were the most frequent offers - excluding equal offers ($10, $10, $10) - in the purely behavioral study we did before starting MRI experiment. We provided proposers with information on the shares of all players before voting irrespective of informational treatment. This was done in order to ensure symmetry between rounds with offers from a human proposer (then he necessarily knows a proposed split) and rounds with computer offers. Effectively our experiment uses 4x2 factorial design with repeated measures. Subjects played VUG repeatedly in 72 rounds: 52 rounds with proposals made by a human proposer and 20 rounds with computer generated proposals. The information conditions were repeated cyclically: Full Info, Power, Me, Other,… The rounds with computer generated offers were dispersed, happening at most once per cycle, and balanced with respect to informational conditions. We let subject playing VUG in the fixed groups of 3 subjects with proposers having a fixed role in the experiment, while the responders were assigned the names of responder 1 or 2 in the unpredictable pattern. The latter was done to obfuscate identity of responders from the perspective of a proposer. After voting took place each subject obtained information about the shares of each voter in a round, irrespectively of informational condition, and whether the proposal was accepted or rejected by majority. No explicit information was given on individual votes by subjects. This was done as we wanted to limit the degree to which the game is perceived by subjects as repeated. The reader can find the exact instructions in the Appendix A.

Table 1: Experimental Treatments.

Treatment	*Message to Responder 1*	*Message to Responder 2*
Full Info	(x, y, z)	(x, y, z)
Power	(x)	(x)
Me	(y)	(z)
Other	(z)	(y)

Condition announced	Waiting for proposal	Evaluation	Deciding	Waiting for voting	Round summary	Mood question	
2000 ms	~10491 ms	3000 ms	~4289 ms	~6175 ms	5000 ms	~3238 ms	time in ms

Figure 1: Timing and Duration of Events within a Round. Average durations of events are reported over all 72 rounds specifically for scanned subjects. Mean round duration was around 34.2 s.

Timing of events is especially important in the analysis of brain data, so we will describe it from the perspective of responders scanned with MRI machine. Each round started with a 2s announcement whether offer is to be made by a human proposer, or generated by a computer. Subsequently, responders were presented with a waiting screen as a proposer was deciding his split of the pie. After this, a proposal was shown for 3 seconds, during which the subject could only evaluate the offer but not act on it. Once the evaluation period was finished the responder had unlimited time to decide whether to accept or reject. If the scanned responder was not the last to vote, the waiting screen followed until all subjects voted. Once this happened the scanned responder was presented with the round summary that included shares of all players, whether majority accepted or rejected the split, and the payoff a player. The summary screen lasted for 5s. Each round ended with a question that "How did that make you feel?" that the subject answered on 11 integer scale with 0 described as "bad", 5 – "neutral," and 10 – "neutral." The question on subjective mood finished each round. We summarize round timing in Figure 1 reporting also average duration of events within a round. We refer the reader to Appendix B for details about fMRI data acquisition, data preprocessing, and modeling brain responses.

4. Behavioral Results

We focus in this study on responders that were scanned using fMRI machine while playing VUG. Consequently, we analyze data from the point of view of a person inside MRI. Due to this it is convenient to adjust the notation denoting with *mri* a share of a responder inside MRI and letting *z* denote a share of responder outside MRI. As before *x* is a share of a proposer. Table 3.2 reports voting patterns of players in response to most frequently encountered offers in all conditions and with proposals made by humans (not a computer). Looking at behavior of MRI responders in condition Full Info they tend to vote to accept almost always offers that give them at least $10 if the proposers are claiming no more than $15. In contrast, looking at offers (x, mri, z) = (15, 5, 10) we notice that MRI responders reject them overwhelmingly, accepting only 8.8% of times. In addition, MRI responders accepted the offer (20, 10, 0) only 70% of times. As expected responders almost always voted to reject offers that gave them $0. In the condition Power the individual acceptance rate by MRI responders is 86% for proposers' demands of $10. If proposers demand more than $10, the acceptance rate drops below 50%. This pattern at treatment Power suggests that responders tried to contain escalating demand of proposers. In turn in condition Me MRI responders vote to accept at least 90% of the times if they are offered $10 or more. Interestingly the acceptance rate of $5 offer is 33.9%, which is more than 8.8%, or acceptance rate of offers (x, mri, z) = (15, 5, 10) in treatment Full Info. In treatment Me responders observed only their own payoff, and consequently their decisions were motivated by inter-personal considerations to a lower degree than in other treatments. Finally in treatment Other, MRI responders seem to vote similarly to three most frequent offers of $0, or $5, or $10 to the other responder. In all cases the individual acceptance rates were slightly above 50% for MRI responders, without significant differences. This suggests that MRI responders weren't concerned with well-being of other responders, and instead they rather randomized votes in treatment Other

Table 2: Responders' Behavior to Most Frequent Offers.

Treatment	*Shares of Players ($)*			*Obs.*	*Voting to Accept (%)*			*RT (ms)*	*Mood*
	x	*MRI*	*other*		*MRI*	*other*	*majority*		
Full Info	**20**	**0**	**10**	25	0	88	88	3391.3	1.8
	15	**0**	**15**	34	2.9	91.2	91.2	3228.1	2.1
	15	**5**	**10**	34	8.8	100	100	3784.1	3.1
	20	**10**	**0**	20	70	15	80	5324.7	5.6
	15	**10**	**5**	29	96.6	55.2	100	4357.9	5.4
	10	**10**	**10**	102	98.0	100	100	3572.3	6.4
	15	**15**	**0**	20	95	0	95	2751.2	7.3
Power	**10**	9.6	10.4	93	86.0	90.3	97.8	4049.5	6.0
	15	7.9	7.1	121	38.0	68.6	82.6	4869.2	4.7
	20	4.4	5.6	44	45.5	40.9	77.3	5081.8	4.1
Me	21.3	**0**	8.7	61	0	82.0	82.0	3196.0	1.8
	20.4	**5**	4.6	56	33.9	44.6	67.9	3246.9	4.4
	18.2	**10**	1.8	148	91.2	18.2	91.9	3084.9	6.2
	14.8	**15**	0.3	20	95	5	95	2630.9	7.7
Other	27.4	2.6	**0**	122	54.1	65.6	82.8	3940.5	2.7
	21.1	3.9	**5**	91	58.2	84.6	92.3	7263.8	2.9
	17.5	2.5	**10**	64	57.8	79.7	92.2	3916.8	2.5

Note: We included above offers with at least 20 observations in a given treatment. The reported offers account for 94.8% of all offers encountered by MRI responder.

to reflect the mixed strategy of a proposer (sometimes taking the whole pie and sometimes leaving some positive shares to responders). If the former were true we would expect larger acceptance rates for payoff of $10 than in case of $0 to the other responder. Overall, the above data suggest that MRI responders primarily maximized their own payoff while engaging in social comparisons with envy and pride. The fairness explanation with aversion to advantageous inequality seems to be inconsistent with patterns observed in treatment Other. In addition, comparing voting patterns of MRI and outside MRI responders, we observe that the former were more aggressive in rejecting comparable offers.

We also analyzed voting behavior of MRI responders running random effects probit regressions, see Table 3. In all models we consider the dependent variable that is equal to 1 if MRI responder accepts and to 0 if he rejects. The regressors include observable shares of players in treatments under study. In particular for treatment Full Info (columns FI in Table 3) we examined three different probit models. One included only the payoff of MRI responder as a regressor and in the other two we added the share of one other player as a regressor. The coefficient on share of MRI responder, or *MRI's share* * $I_{[FI]}$ is highly

Table 3: Random Effects Probit Models of Responders' Acceptance

Variable	*Responder's Vote to Accept (1 - Accept, 0 – Reject) across Treatments (T)*					
	FI	*FI*	*FI*	*FI&Power*	*FI&Me*	*FI&Other*
MRI's share * $I_{[T]}$	0.489***	0.512***	0.721***		0.547***	
	(0.071)	(-0.101)	-0.151		(0.054)	
Proposer's share* $I_{[T]}$		-0.210**		-0.203***		
		(-0.071)		(0.021)		
Other's share * $I_{[T]}$			0.210**			-0.045***
			-0.071			(0.012)
Proposer's share * $I_{[Power]}$				-0.004		
				(0.008)		
MRI's share * $I_{[Me]}$					-0.003	
					(0.022)	
Other's share * $I_{[Other]}$						-0.024
						(0.016)
Constant	-3.211***	-0.119	-6.414***	3.256***	-3.62***	0.550***
	(0.564)	-1.04	-1.51	(0.346)	(0.467)	(0.145)
Obs.	286	286	286	572	572	572
No subject	22	22	22	22	22	22
Log-likelihood	-61.92	-53.55	-53.55	-287.2	-130.3	-361.8

Note: Standard errors are reported in parenthesis. P values convention: *** p<0.001, ** p<0.05, * p<0.1.

significant ($p<0.001$) and positive in all three cases. The additional effect of shares of other players, either a proposer or other responder, is relatively smaller, but significant ($p<0.05$). However, we have to be careful in the interpretation of coefficients on other players' shares: *Proposer's share* $* I_{[FI]}$ and *Other's share* $* I_{[FI]}$. The reason is collinearity problem that arises due to the fact shares of players always summed up to \$30. In this light, the rather robust conclusion is that the acceptance probability by MRI responder is negatively influenced by proposer's demand as the coefficients on *MRI's share* $* I_{[FI]}$ are similar in probit models that include *Proposer's share* $* I_{[FI]}$ as regressors or not. This conclusion is strengthened once we consider a probit model for treatments Full Info and Power together with two regressors *Proposer's share* $* I_{[FI\&Power]}$ and *Proposer's share* $* I_{[Power]}$. We obtain the result that the impact of proposers share in on MRI responder's acceptance does not differ between treatments Full Info and Power, notice that coefficient on *Proposer's share* $* I_{[FI\&Power]}$ is similar to the one obtained in treatment Full Info, or *Proposer's share* $* I_{[FI]}$. Further, the coefficient on *Proposer's share* $* I_{[Power]}$, which is to measure the additional impact on MRI responder's decision to accept in treatment Power, where only proposer's payoff is observable in contrast to treatment Full Info, is not significant. We can conclude that proposer's share impacts negatively acceptance probability of MRI responders. In contrast, the impact of other responder's share is less clear at first. Note that in the probit model for Full Info treatment, inclusion of *Other's share* $* I_{[FI]}$ as a regressor causes significant increase in coefficient on MRI's share $* I_{[FI]}$ as compared to other probit models. The significance of coefficient on *Other's share* $* I_{[FI]}$ may be the result of colinearity problem. This interpretation is more likely given that analyzing treatment Other the impact of *Other's share* $* I_{[FI\&Other]}$ on MRI responder's acceptance probability doesn't differ between treatments Full Info and Other. In other words a coefficient on *Other's share* $* I_{[Other]}$ is close to zero and not significant. At the same time coefficient on *Other's share* $* I_{[FI\&Other]}$ is negative and significant. This suggests that the negative impact of other responder's share on acceptance probability of MRI responders. Summarizing, the analysis suggests that MRI responder is primarily selfish and affected negatively by payoffs of other players. Moreover, MRI responders are affected negatively more by proposer's than by responder's payoff.

5. Neural Results

The analysis of neural data collected in the experiment involves some challenges that are consequences of constraints one encounters in MRI experiments as opposed to purely behavioral experiments. Due to the cost of such undertakings our MRI experiment involves small sample, this is not different from others. However we differ from others, including from Sanfey et al (2003) that is the most lcolesely related paper, in that we didn't use deception in order to alleviate some practical constraints related to MRI experiments. On the one hand provides more credibility to our results, and on the other has a negative side effect that we had no control over offers made by human proposers in our study. As the result our study is unbalanced with respect to offers: different responders faced different sets of offers. After conducting extensive exploratory analysis of the data we decided to focus in this paper on the question what are neural underpinnings of rejections of positive amounts of money in our 3-person ultimatum game with voting in Full Info treatment. This alone turned out to be sufficiently daunting and interesting task to deserve a separate paper that follows below. Other interesting issues related to the MRI experiment under consideration, for example analysis of neural data between conditions or of social comparisons, are primarily left for separate investigation.

5.1. Contrast Analysis

We start analysis with comparison of neural responses to splits (x, mri, z) = (*, 0, *) and (*, 5, *) made by human proposers in treatment Full Information. The former offer type corresponds mainly (see Table 2 above) to offers like (20, 0, 10) or (15, 0, 15), and the latter type offer like (15, 5, 10) and to lower extent (20, 5, 5). There were 11 scanned responders in our sample, who faced both types of offers. We know from the Table 2 that scanned responders were essentially rejecting both. However, it seems reasonable to think that the neural mechanisms behind rejections to these two types of offers are different. From the perspective of a responder voting to reject the offer of $0 to himself seems natural, by contrast rejection of the offer of $5 seems less so, why giving up money? To get insight into the neural processes behind responses to these two types of offers we focused on the evaluation period. We used the study design in which specifically we modeled each offer with separate dummy (stick) predictors twice, in the beginning and the end of the evaluation period. For example, for a given offer we had one dummy predictor in the initial 500ms of evaluation period and the second one in the last 500ms. The inter-time[2] of 2000ms was modeled as offer-invariant dummy and all other events were modeled with separate dummy predictors. We chose such modeling approach motivated by Sanfey et al (2003), who suggested that unfair but positive offers generate in responders a cognitive conflict between two opposing processes: cognitive and emotional ones. This account implicates anterior cingulate cortex (ACC) in detecting the conflict between right dorso-lateral prefrontal cortex (right DLPFC) – cognitive process, and anterior insula – emotional process. Specifically, Sanfey et al (2003) suggested that right DLPFC is related to goal maintenance of maximizing monetary gain, while increased activity in bilateral anterior insula reflects an emotional reaction to unfairness. However, given this interpretation it remains unclear why DLPFC, which is associated with executive control rather than reward processing, is implicated at the first place. More precisely, one could expect that decision conflict comes from mismatch between reward valuation and emotional response to unfair offers, and then ACC detects it, and further regions are employed to resolve it, including right

[2] Note that choosing the inter-time of 2000ms between predictors can be partially justified that repetition time (TR), period within which whole-brain data were collected anew, was also 2000ms. In principle one could consider slightly different inter-time values as compared to the one equal exactly to TR. However, there is no obvious way to optimize the inter-time view from the point of view of statistical analysis.

DLPFC. Our statistical model design, described above, was specially devised to explore the hypothesis of such valuation vs. emotion conflict, and its subsequent resolution. It is reasonable to expect that some basic valuation and emotional processes are automatic and instantaneous once stimulus is encountered. Similarly, once this happens checking whether valuation and emotion are in conflict should also be rather automatic, together with some initial steps that aim at resolving a conflict, if such conflict is present at all. In contrast, it is equally reasonable to expect that there are other brain processes related to decision-making that are employed at discretion of decision-maker. We are not able to analyze such discretionary processes given current technology and methods of analysis. However, we should be able to say more about processes that are automatic as they ensue in short period of time once the stimulus, or as in our case offers to responders, are prese**nte**d

Our results are illuminating. Looking at the contrast of offer (*, 5, *) versus (*, 0, *) as modeled by dummy predictors in the initial evaluation phase we get increased activations in nucleus accumbens (NAcc) and amygdala, see Table 4 as wells as panels A and B in Figure 2. On the one hand, NAcc activity was shown in many studies to correlate positively with reward value of stimuli, including our study reported in Grygolec, Coricelli, Rustichini (2008) where activity in NAcc tracked experienced reward. On the other hand increased activity of amygdala was associated with expression of anger and fear. In the light of what we know of function of NAcc and amygdala our data provide evidence that offers like (*, 5, *) in treatment Full Info lead to conflict between nucleus accumbens (NAcc) that records positive value of monetary gain

Table 4: Brain Areas Showing increased Activity in the Contrast of Unfair Offer vs. Null Offers in Full Information Treatment

Evaluation Phase	*ROI*	*Side*	*Tal X*	*Tal Y*	*Tal Z*	*Voxels*	*Max t*
Initial	NAcc	right	12	5	-8	69	4.9
	Amygdala	right	30	-5	-16	75	4.8
	Amygdala	left	-18	-5	-16	708	6.9
Late	Insula	right	33	14	10	112	5.9
	Insula	left	-31	19	7	33	4.5
	DLPFC	right	42	45	19	86	6.6
	OFC	n/a	0	50	-7	25	4.2
	ACC	n/a	5	22	35	182	4.9

Note: Statistical threshold: p<0.005 uncorrected, cluster size > 5 voxels. Labels based on initial Talaraich coordinates (reported) assigned to areas of interest.

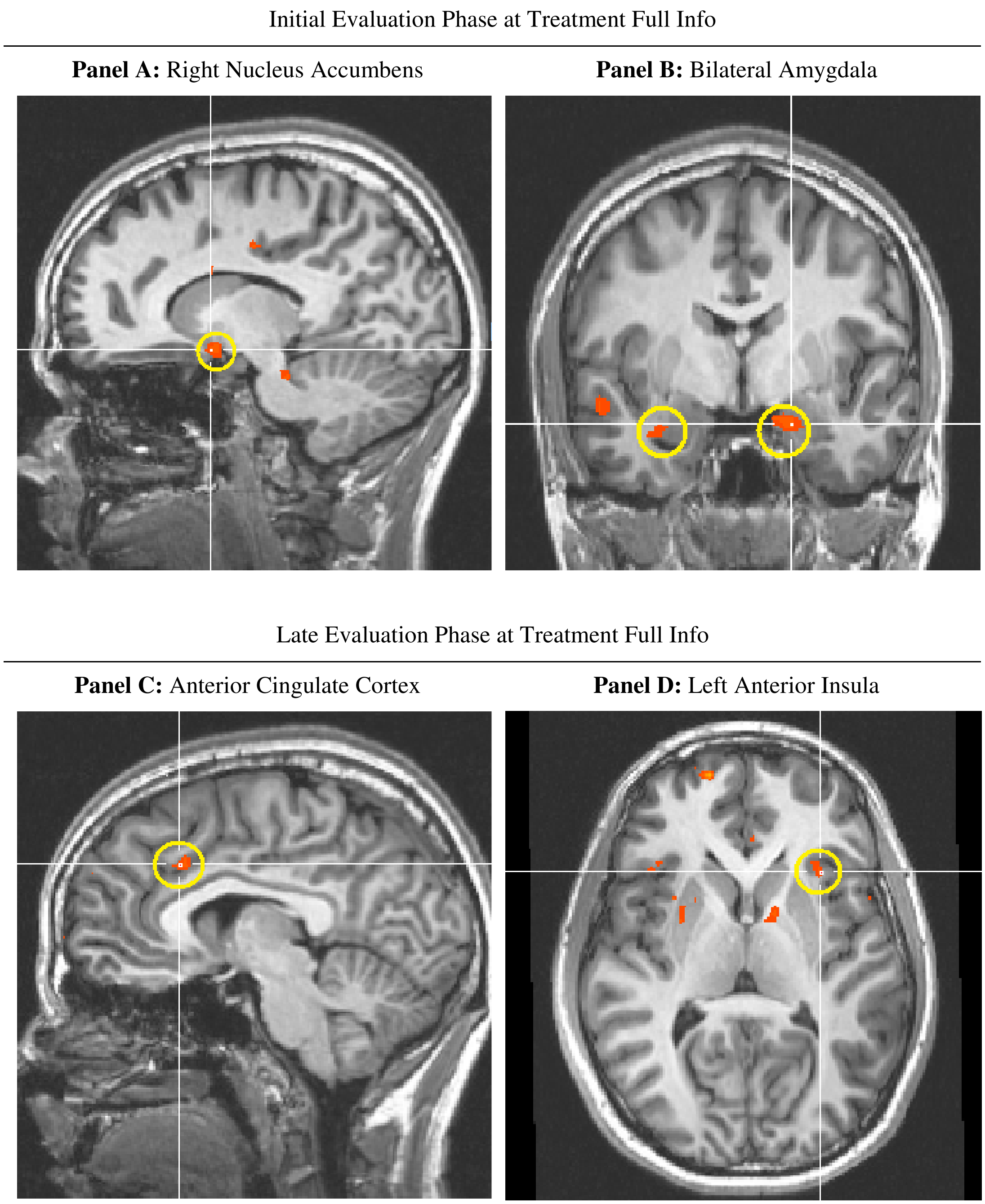

Figure 2: Regions with Increased Activity to Low Positive Offers. All panels refer to contrasts with the responder in MRI receiving \$5 versus \$0 at treatment Full Info with proposals made by a human. Upper panels A and B show increased activity at early evaluation stage in right nucleus accumbens and bilateral amygdale, respectively. In turn, lower panels C and D show increased activity at late evaluation phase in anterior cingulate cortex and left anterior insula, respectively.

and amygdala that conveys negative emotional response to the very same offer. This conflict has to be resolved one way or another before the actual decision is taken, but this may require that the brain employs additional regions in analyzing the stimuli. To identify brain regions that may be involved in conflict resolution we investigated the contrast of offer (*, 5, *) versus (*, 0, *) as modeled by dummy predictors in the late phase of evaluation period, see Table 4 as wells as panels C and D in Figure 2. We observed increased activity in ACC and right DLPFC, regions that were associated in the literature by Sanfey et al (2003) and Pochon et al (2008) with conflict monitoring, and executive control (or goal maintenance), respectively. We found also in the same contrast with late predictors increased activation in anterior insula bilaterally, and medial OFC. The increased activity in the former region is usually associated with disgust or pain. Reconciling all these findings together we suggest that increased activity in ACC detects the conflict between positive relative valuation as captured by nucleus accumbens and negative emotion reflected by amygdala in response to the offer (*, 5, *) - as benchmarked to the offer (*, 0, *).

In the light of the detected conflict yet another region associated with negative emotion is being used before decision is taken: bilateral anterior insula. As insula activates usually to negative stimuli it looks as the offer at the center of conflict is reevaluated only from negative side. Further, we suggest that right DLPFC is involved in integrating different conflicting signals, both from NAcc and amygdala in the initial stage of offer evaluation, and from anterior insula in the later stage. This last suggestion is based on the results of Knoch et al (2006, 2007). They used brain stimulation techniques to suppress brain activity in right DLPFC and as the result observed increased acceptance rates of unfair offers in 2-person ultimatum game as compared to sham treatment. However, in the earlier study of Sanfey et al (2003) the activity in right DLPFC showed increased activity during evaluation of unfair as compared to fair offers, but crucially ir didn't correlate with rejections of unfair offers in 2-person ultimatum game. One possible reason for these allegedly incompatible findings could be that the right DLPFC integrates different conflicting signals from nucleus accumbens, amygdala and insula that are related to offer evaluation. We explore this possibility below.

5.2 Region of Interest Analysis

We performed region of interest (ROI) analysis in the areas implicated in decision processes behind MRI responders' responses to offer (x, mri, z) = (*, 5, *). The main motivation for performing ROI analysis is that it allows to account for subject heterogeneity and to test our results obtained above at between-subject level. Specifically, we used model design that included predictors for each distinct offer type that was made at least 20 times by all proposers[3]. As before, we created for each offer two predictors one corresponding to the initial and one to the late evaluation phase. Figure 3 shows activity in nucleus accumbens with offers lexicographically ordered: first increasing MRI responder's share, second increasing other responder's share. In general we observe the trend for activity in nucleus accumbens to increase with MRI responders' payoff. We examined a robust[4] linear regression with lexicographic rank of offer as independent variable and percentage BOLD change in right nucleus accumbens (NAcc) only to obtain a positive coefficient on rank: 0.873 (z 1.82, p 0.072). The similar ROI analysis for right amygdala shows a different pattern, see Figure 4. Notably, the activity at this region achieves local peak at offer (x, mri, z) = (15, 5, 10). At the same time right amygdala exhibits relatively lower activity in response to offers that give MRI responders $0 (z 2.154, p 0.031, Mann-Whitney rank sum test, two-sided) or $10 with proposer's advantage at the same time (z 1.663, p 0.0963, Mann-Whitney rank sum test, two-sided). Taken together the ROI analysis provides further evidence that the offer (x, mri, z) = (15, 5, 10) leads to conflicting signals: from nucleus accumbens (measuring positive reward value) and amygdala (exhibiting negative emotional response).

To check whether anterior cingulate cortex is involved in detecting the aforementioned conflict we performed regression whether percentage BOLD change to the offer (x, mri, z) = (15, 5, 10) in right nucleus accumbens and right amygdala can predict percentage BOLD change in anterior cingulated cortex, see Table 5. Taken alone activity in right amygdala fails to predict activity in anterior cingulated cortex, model 1 in Table 5. However, once we add activity in right nucleus accumbens or interaction term as independent variables, activity in amygdala predicts positively activity in anterior cingulated cortex, see models 2-4 in Table 5. At the same time activity in right nucleus accumbens or interaction term correlates negatively with activity in anterior cingulate cortex. Note, the positive coefficient on percentage BOLD change in right nucleus accumbens in model 3, that includes all independent variables considered, is not significant (p>0.1). We dropped this predictor in model 4 of anterior cingulated cortex activity only to obtain positive coefficient on activity in right amygdala and negative coefficient on interaction term, both being highly significant (p<0.001). Notably, R-squared for model 4 stands at 0.87, which suggests that it has substantial explanatory power.

---

[3] The offers used are reported in Table 2. They account for 94.8% of all offers made by human proposers. The unusual offers were modeled with separate predictors and were of no interest. Distinct computer offers were modeled with separate predictors.

[4] Using robust regression is justified by outlier problem. It is conservative approach here since standard linear regression gives also a positive coefficient on rank 1.731, and even more significant (z 2.65, p 0.01).

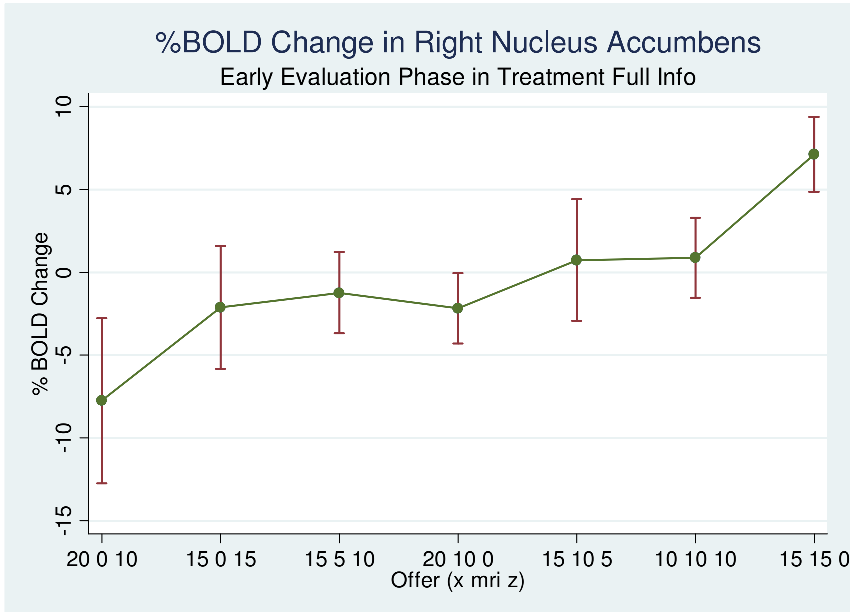

Figure 3: Percentage BOLD Change in Right Nucleus Accumbens. It applies to early evaluation phase in treatment Full Info for different offers of human proposers. For each offer whisker plot indicates mean ± standard error.

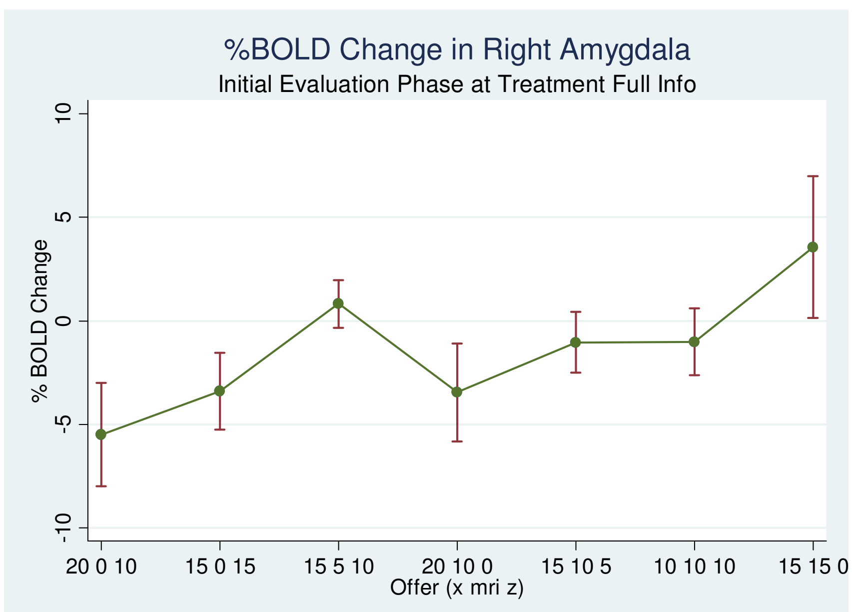

Figure 4: Percentage BOLD Change in Right Amygdala. It applies to initial evaluation phase in treatment Full Info for different offers of human proposers. For each offer whisker plot indicates mean ± standard error.

Table 5: Anterior Cingulate Cortex Detecting Conflict between Reward Value and Emotional Response to Low Positive Offer: Between-Subject Analysis

Variables	Anterior Cingulate Cortex %BOLDΔ			
	Model 1	Model 2	Model 3	Model 4
right Amygdala %BOLDΔ	15.380	14.530*	47.667**	33.120***
	(10.786)	(7.174)	(9.089)	(5.527)
right Nucleus Accumbens %BOLDΔ		-32.964**	31.546	
		(9.890)	(16.817)	
Interaction term = (Amygdala*NAcc)			-57.842**	-32.602***
			(14.244)	(5.450)
Constant	-7.977	14.615	-24.003*	-4.824
	(10.288)	(9.629)	(10.922)	(4.480)
Observations	10	10	10	10
Log-likelihood	-39.96	-35.21	-28.60	-30.91
R-squared	0.203	0.692	0.918	0.870

Note: *** p<0.001, ** p<0.05, * p<0.1 Standard errors in parentheses. To ensure comparability of percentage BOLD changes between subjects we performed subject by subject normalization to [0,1] interval based on maximum and minimum percentage BOLD changes observed to 6 most frequent offers (see Table 2) in treatment Full Info. The results obtained are robust to normalization if one outlier is dropped from non-normalized data.

The ROI analysis above, taking into account subject heterogeneity, strengthened the case for the interpretation that low positive offers like (x, mri, z) = (15, 5, 10) evoke in MRI responders the conflict between positive reward value and negative emotional response, and that it is being detected by anterior cingulate cortex. Now we want to investigate in similar manner, taking into account subject heterogeneity, how this conflict is being resolved. The key brain areas possibly involved, as discussed above, are bilateral insula and right DLPFC. Figure 5 shows that percentage BOLD change in right[5] insula increases in response to offer (x, mri, z) = (15, 5, 10) as compared to offers in which MRI responders receive $0 (Mann-Whitney rank sum test, z 1.701, p 0.089), as well as compared to offer (x, mri, z) = (15, 10, 5) (Mann-Whitney rank sum test, z 1.861, p 0.0628). The former finding is startling: the offer of $0 to MRI responder doesn't elicit negative emotional response as measured by activity in insula in contrast to the offer of $5. In our view this finding calls for reinterpretation of rejections of low positive offers in bargaining experiments in general. Put simply, responders are concerned with their image and they don't

[5] The similar results apply to the left insula.

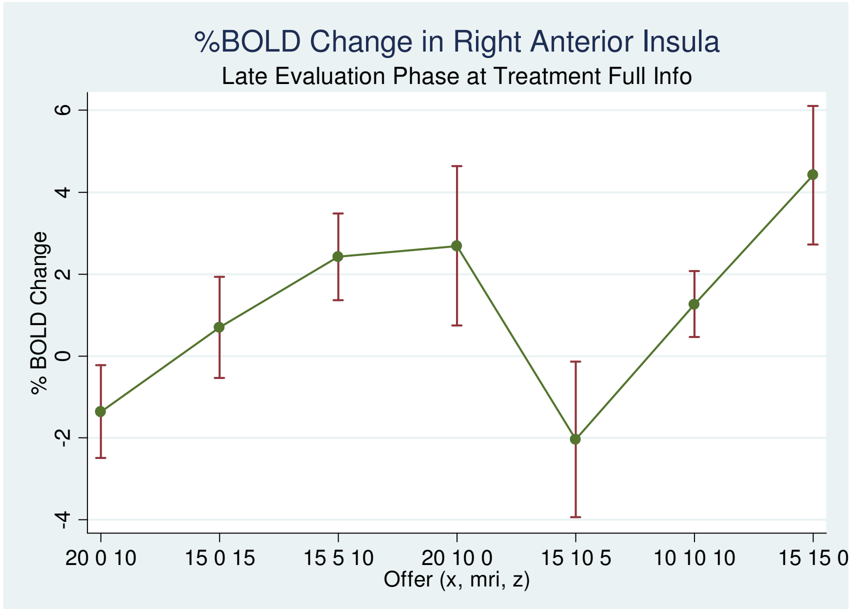

Figure 5: Percentage BOLD Change in Right Anterior Insula. It applies to early evaluation phase in treatment Full Info for different offers of human proposers. For each offer whisker plot indicates mean ± standard error.

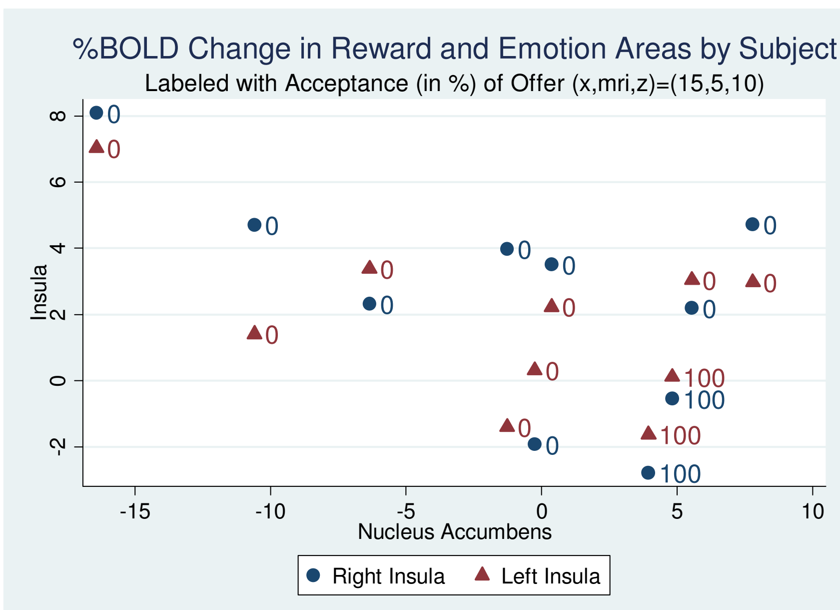

Figure 6: Activity in Reward and Emotion Areas Predicts Rejections. %BOLD change is shown in right nucleus accumbens versus right and separately left insula to offer (x, mri, z) = (15, 5, 10) between subjects. Each subject corresponds to a distinct value of percentage BOLD change in right nucleus accumbens. Two subjects with both high activity in nucleus accumbens and low in bilateral insula always accept such offers. In contrast, other subjects reject such offers.

want to be perceived as weak types. At the same time offers in which MRI responder has advantage over the other responder also are associated with increase in percentage BOLD change in right insula. However, standard errors also increase, and in addition these apparent negative emotional reactions are not correlated with increased rejections of the corresponding offers. We view this pattern to reflect the initial uncertainty of MRI responder as to whether they or the other responder received higher payoff once the offer was being displayed[6]. This interpretation seems more plausible than the alternative explanation that MRI responders were reacting emotionally to the fate of the other responder receiving lower payoff. If it weren't the case, why no negative emotions ensued in MRI responders when they were treated unfairly. Yet, another possibility is that the negative emotional reaction of MRI responders to offers with the other responder receiving the lowest payoff is due to the guilt of being an accomplice in proposer's scheme to dominate.

We suggest, in accordance with the presented evidence, that the offer (x, mri, z) = (15, 10, 5) evoking conflict between positive reward value and negative emotions is reevaluated only from the viewpoint of negative emotional significance (bilateral insula) at the late evaluation phase. In this light we would expect \behavior of subjects in terms of either accepting or rejecting the offer to be related to subjective value measures taken in the initial evaluation phase in right nucleus accumbens and reevaluated negative emotion as reflected by activity in bilateral anterior insula. Figure 6 shows scatter plot by subject of percentage BOLD change in right nucleus accumbens versus those in right and left insula (treated separately). Note that each subject corresponds to distinct activity of right nucleus accumbens and labels specify acceptance rates of the corresponding subjects to the offer (x, mri, z) = (15, 10, 5). Only the subjects with high activity in nucleus accumbens (high reward value) and low activity in bilateral insula (low negative emotion) accept such offers. Other subjects in contrast always reject. Knoch et al (2006, 2007) showed that right DLPFC seems critical to rejections of positive but unfair offers in 2-person ultimatum game, and if its activity is suppressed the acceptance rates increase significantly. In line with this finding we find that right anterior insula (in contrast to right nucleus accumbens) predicts activity in right DLPFC, see Table 6. At the same time we don't find correlation of DLPFC activity with rejection rate. This may be due to the fact that right DLPFC, traditionally connected in the literature to executive control, may be incorporating neural signals into decision that are in general unrelated to pure subjective value of stimuli (as measured by nucleus accumbens). For example in the current context negative emotional response is suggested to be such a signal, but there is little a priori reason to believe that it alone exhausts the list of signals that right DLPFC may incorporate into decision.[7]

[6] We randomized the position of the share going to MRI responder from trial to trial. We wanted to remind subjects that the identity of responders (Voter 2 or Voter 3) was changing from to trial to trial.

[7] For the sake of example consider that a subject, taking part in the experiment, decided to accept every offer no matter what because he didn't believe that offers were being made by real subjects.

Table 6: Right Dorso-lateral Prefrontal Cortex Mediating Rejections of Offers due to High Negative Emotion as measured by Right Insula

Variables	Right DLPFC %BOLDΔ	
	Model 1	Model 2
right Insula %BOLDΔ	1.172**	1.128**
	(0.358)	(0.411)
right Nucleus Accumbens %BOLDΔ		-0.080
		(0.280)
Constant	-0.154	-0.066
	(0.294)	(0.440)
Observations	10	10
Log-likelihood	1.147	1.204
R-squared	0.573	0.578

Note: *** p<0.001, ** p<0.05, * p<0.1 Standard errors in parentheses. To ensure comparability of percentage BOLD changes between subjects we performed subject by subject normalization to [0,1] interval based on maximum and minimum percentage BOLD changes observed to 6 most frequent offers (see Table 2) in treatment Full Info. The results obtained are robust to normalization.

5. Conclusions

The popular view explaining paradoxical behavior of experimental subjects in 2-person ultimatum game - rejecting positive amounts of money – implicates preference for fairness at the center stage. This view was generalized by theories of fairness of Fehr, Schmidt (1999), or Bolton, Ockenfels (2000) and points both to aversion to disadvantageous and advantageous inequality (strong fairness hypothesis). Our results discussed in Grygolec, Coricelli, Rustichini (2008) are consistent with the former (*envy* to avoid multiple negations) on the one hand and on the other hand inconsistent with the latter. The evidence presented pointed to pride rather than the opposite of pride - aversion to advantageous inequality. We found neuroeconomic evidence for both envy and pride exhibited by experimental subject while evaluating outcomes of simple decisions in social environment. Most importantly, the neural subjective value of outcomes, as measured by activity in nucleus accumbens, was sensitive to social context, exhibiting both envy and pride. This effect was stronger if a subject had relatively active as opposed to passive role in generating the difference in outcomes between people. In this paper we discussed the results of fMRI experiment of responders playing 3-person ultimatum game while being scanned with MRI machine. We focused on neural mechanisms associated with rejections of positive amounts of money in 3-person ultimatum game. Crucially, previous research in neuro-economics implicated negative emotion to such rejections, i.e. activity in bilateral insula correlated with rejections in Sanfey et al (2003). The interpretation of this result was that negative emotion arises in responders to the perceived unfairness of offers in which a proposer demands substantial advantage in the ultimatum game. Our results point to fundamental reassessment of this view. Notably, the negative emotional response as measured by activity in anterior insula bilaterally and bilateral amygdala is larger to offers in which an MRI responder is offered $5 rather than $0. Consequently, negative emotion expressed by responders in response to offers with propoposer's advantage can NOT alone be interpreted as being driven by unfairness of such offers. After all offers like (x, mri, z) = (20, 0, 10) or (15, 0, 15) should be considered as more unfair from the viewpoint of MRI responder than the offer (15, 5, 10) according to any reasonable theory of fairness. Rather we suggest that the main factor behind negative emotional responses of MRI responders is related to what signal acceptance or rejection of low but positive offer may send to others. Accepting the offer with minimal positive amount of money like $5 in the offer (15, 5, 10) signals, if anything, the weak type of the responder. Furthermore, in Grygolec, Coricelli, Rustichini (2009) we implicated right nucleus accumbens, the area shown in the literature to track the subjective reward value of stimuli, to be involved in social comparisons exhibiting envy and pride effects. We have shown in this paper that nucleus accumbens seems to track subjective value of offers[8] in 3-person ultimatum game with voting. This leaves open the possibility that social comparisons exert the imprint on responders' decisions via the route of nucleus accumbens. We leave the investigation of social comparisons versus strong fairness hypotheses in the context of ultimatum bargaining for the separate treatment. In this context it is worth considering whether one should incorporate randomized pie size into experimental design in order to be able to indentify inter-personal preferences with weaker assumptions.

This paper sheds also more light on the neural basis of decision conflict leading to rejections of positive but low amount of money. Building on the contributions of Sanfey et al (2003) and Knoch et al (2006, 2007) we separated the decision conflict into two phases during offer evaluation: the early and late phase. During the former phase the offer is being evaluated by nucleus accumbens (measuring reward value) and amygdala (measuring negative emotional response). If conflict arises between these two regions involved in offer evaluation, anterior cingulate cortex detects it subsequently in the late stage. Further, it seems as offers leading to decision conflict are re-evaluated by bilateral anterior insula for its negative emotional significance in the late stage. Finally, the right dorso-lateral prefrontal cortex possibly mediates into decisions the negative emotion as captured by right anterior insula among other factors.

[8] More precisely, offers ordered lexicographically from the view point of MRI responder with his payoff first and the payoff of the second responder second.

References

[1] Bolton, G. and A. Ockenfels, (2000): "Strategy and Equity: A Theory of Equity, Reciprocity, and Competition," American Economic Review, Vol. 90. 166-193.

[2] Burnham, T. C. (2007): "High-testosterone Men Reject Low Ultimatum Game Offers," Proceedings of the Royal Society B, Vol. 274, 2327-2330.

[3] Camerer, C. F. (2003): Behavioural Game Theory. Experiments in Strategic Interaction, Princeton: Princeton University Press.

[4] Emanuele, E., N. Brondino, M. Bertona, S. Re, D. Geroldi (2008): "Relationship between Platelet Serotonin Content and Rejections of Unfair Offers in the Ultimatum Game, Neuroscience Letters, Vol. 437, 158-161.

[5] Grygolec, J., G. Coricelli, A. Rustichini (2008): "A Neuroeconomic Study of Social Obervability and Personal resobsibility in Decision Making: An fMRI Experiment," unpublished manuscript.

[6] Güth, W., R. Schmittberger, and B. Schwarze, (1982): "An Experimental Analysis of Ultimatum Bargaining," Journal of Economic Behavior and Organization, Vol. 3, 367–388.

[7] Güth, W., and E. van Damme (1998): "Information, Strategic Behavior, and Fairness in Ultimatum Bargaining: An Experimental Study," Journal of Mathematical Psychology, Vol. 42, 227-247.

[8] Fehr, E. and K. M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation, Quarterly Journal of Economics, Vol. 114, 817-868.

[9] Hampton, A. N., R. Adolphs, J. M. Tyszka, and J. P. O'Doherty (2007): "Contributions of the Amygdala to Reward Expectancy and Choice Signals in Human Prefrontal Cortex," Neuron, Vol. 55, 545-555.

[10] Knoch, D., M. A. Nitsche, U. Fischbacher, C. Eisenegger, A. Pascual-Leone, and E. Fehr (2007): "Studying the Neurobiology of Social Interaction with Transcranial Direct Current Stimulation – The Example of Punishing Unfairness," Cerebral Cortex, Vol. 18, 1987-1990.

[11] Knoch, D., A. Pascual-Leone, K. Meyer, V. Treyer, E. Fehr (2006): "Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex," Science, Vol. 314, 829-832.

[12] Koenigs, M. and D. Tranel (2007): "Irrational Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game," Journal of Neuroscience, Vol. 27, 951-956.

[13] Pochon, J.-P., J. R. Riis, A. G. Sanfey, L. E. Nystrom, and J. D. Cohen (2008): "Functional Imaging of Decision Conflict," Journal of Neuroscience, Vol. 28, 3468-3473

[14] Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen (2003): "The Neural Basis of Economic Decision-Making in the Ultimatum Game," Science, Vol. 300, 1755-1758.

[15] Wallace, B., D. Cesarini, P. Lichteinstein, and M. Johannesson (2007): "Heritability of Ultimatum Game Responder Behavior," Proceedings of National Academy of Sciences, Vol. 104, 15631-15634.

[16] van't Wout, M., R. S. Kahn, A. G. Sanfey, and A. Aleman (2005): "Repetitive Transcranial Magnetic Stimulation over the Right Dorsolateral Preforntal Cortex Affects Strategic Decision-making," Neuroreport, Vol. 16, 1849-1852.

Appendix A: Instructions

In the beginning of the experiment subjects were placed in separate rooms and given a paper version of PowerPoint presentation containing instructions. We asked subjects to read these instructions privately. Subjects were allowed to ask questions that were answered by the experimenter. We report below the content of PowerPoint presentation used to instruct subjects. Slide titles are double-quoted followed by ordered bullets in separate sentences.

Slide 1: "Instructions" Please read carefully.

Slide 2: "Instructions" You will play a game repeatedly. The game is among 3 people including you. You will play with the same two people in all rounds, but you will never know who they are

Slide 3: "The Game" You will assume one of three roles: Voter 1, Voter 2, or Voter 3. You will know your role at the beginning of each round. If you are voter 1 you remain voter 1 in all rounds. The identities of Voters 2 and 3 will change randomly each round between the other 2 persons.

Slide 4: "Voting" In each round Voter 1 proposes a division of $30 among the three voters. Each voter will vote to accept or reject a proposal. If at least 2 voters accept, then the proposed split is carried out. Otherwise each one receives $0.

Slide 5: "Split by Voter 1 or Computer" Sometimes **Voter 1** and sometimes a **computer** proposes a split of $30. Everyone will know who makes a proposal. Voter 1 is asked to make a proposal so that shares of voters are multiples of $5

Slide 6: "What do voter 2 and 3 know before voting?" There are 4 conditions: Condition 1: Each voter sees ***the share of each voter***. Conditions 2-4: Each voter sees ***the share of only one voter…***

Slide 7: "What do voter 2 and 3 know before voting?" Condition 2: they know the share of voter 1. Condition 3: they know their own share. Condition 4: they know the share of the other voter.

Slide 8: "After voting" After voting everyone learns: the whole proposal; whether the proposal was accepted (at least 2 votes for) or rejected; his/her payoff.

Slide 9: "Payment" In the end a computer will draw one period for which we will pay each voter.

Appendix B: Materials and Methods

Data Acquisition

We scanned 22 healthy male subjects. Brain data were acquired using Siemens 3T TIM Trio scanner at the Center for Magnetic Resonance Research at the University of Minnesota. In each scanning session we collected first anatomical and then functional data. The anatomical data were obtained using T1-weighted fast 3D gradient echo pulse sequence with the parameters: voxel size 1×1×1 mm, 176 slices per slab, slice thickness 1 mm, data matrix 224×256, repetition time (TR) 2600ms, echo time (TE) 3.02ms. The functional data were collected with echo planar T2*-weighted sequence with voxel size 3×3×3, slice thickness of 2 mm with inter-slice gap of 1 mm, 36 slices per image, data matix 64×64, TR 2000 ms, TE 25 ms.

Preprocessing

fMRI data were preprocessed and analyzed using BrainVoyager QX version 1.96 (Brain Innovation B.V., Maastricht, the Netherlands). We first applied iso-voxel transformation to anatomical data to obtain data in square data matrix 256 ×256, and then transformed the data into Talaraich space. We used standard procedure for Talaraich transformation with first rotating the cerebrum into anterior commissure – posterior commissure (AC-PC) plane, and second identifying borders of the cerebrum so that the size of the brain was fitted into Talaraich space. We preprocessed functional data performing slice scan time correction, 3D movement correction relative to the first volume using tri-linear estimation and interpolation, removal of linear trend together with low frequency non-linear trends using high-pass filter. Next, we co-registered functional with anatomical data to obtain Talaraich referenced voxel time courses, to which we applied spatial smoothing using Gaussian filter of 7 mm.

Modeling Brain Responses

We used box car predictors to model events of no interest. To model events of interest we used stick predictors with categorical variables being the product of indicator function of condition (information about offer, and whether offers was generated by human or a computer) and value of stimulus/decision. We explain predictors of interest in more detail in the main text as we present results.. We convolved predictors with a standard two gamma hemodynamic response function (hrf) that models blood oxygenation level dependent signal measured by fMRI technique.