International Journal of Game Theory

FORTHCOMING 2016

Strategic Risk and Response Time Across Games*

Pablo Brañas-Garza† Debrah Meloso‡ Luis Miller§

June 17, 2015

Abstract

Experimental data for two types of bargaining games are used to study the role of strategic risk in the decision making process that takes place when subjects play a game only once. The bargaining games are the Ultimatum Game (UG) and the Yes-or-No Game (YNG). Strategic risk in a game stems from the effect on one player's payoff of the behavior of other players. In the UG this risk is high, while it is nearly absent in the YNG. In studying the decision making process of subjects we use the time elapsed before a choice is made (response time) as a proxy for amount of thought or introspection. We find that response times are on average larger in the UG than in the YNG, indicating a positive correlation between strategic risk and introspection. In both games the behavior of subjects with large response times is more dispersed than that of subjects with small response times. In the UG larger response time is associated with less generous and thus riskier behavior, while it is associated to more generous behavior in the YNG.

Keywords: Response Time, Ultimatum Game, Yes-or-No Game, Strategic Risk.

1 Introduction

We use *Response Time* (RT) and behavioral data from two games recreated in the laboratory to test the hypothesis that individuals use introspection, reflecting about the strategic risk of the situation they face, when confronted with a new strategic situation. The situations we consider are the *Ultimatum Game* (UG) and the *Yes-or-No Game* (YNG). Both

*The authors gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation (grants ECO2012-30626, CSD2010-00034, and ECO2013-44879-R), the Government of Andalusia Project for Excellence in Research (grant P12-SEJ-1436), Basque Government (research group IT-783-13), and the European Research Council (advanced grant BRSCDP-TEA). We thank Praveen Kujal, Jaromir Kovarik, Ismael Rodríguez-Lara, and the audience at the 4th annual meeting of SEET in Tenerife for useful comments.

†Middlesex University Business School, Hendon Campus, The Burroughs, London NW4 4BT, UK

‡ESC-Rennes School of Business **Corresponding author**. debrah.meloso@esc-rennes.com

§School of Economics and Business, University of the Basque Country (UPV/EHU), Av. Lehendakari Aguirre 83, 48015 Bilbao, Spain

are two-player games, one player being the *proposer*, and the other the *responder*, and in both games the players must agree about the division of a valuable object. The proposer in either game makes a proposal about the division of the object that is implemented if the responder accepts the division, while both players receive a payoff of 0 if the responder rejects. The games differ in that the responder in the YNG can either accept all offers or reject all offers. In the UG the responder can condition acceptance on the value of the offer made by the proposer.[1] The strategy sets of proposers are identical across the two games, while the strategy sets of responders differ in a way that significantly lessens the value to the proposer of predicting the responder's behavior.

In our experiment we have each subject participate only once in only one game, holding only one of the two possible roles. We do this because our questions regard *one-shot* games. The study of one-shot games is important because many of the applications of game theory in the social sciences involve situations that the relevant actors face only once or seldom. Examples include negotiations over work contracts, the buying of property, voting on committees under different rules, and many decisions related to marriage and parenting. A crucial characteristic of one-shot games is that one cannot reasonably assume that agents *know* how other players will behave. Knowing how others play is fundamental to the notion of Nash equilibrium: each agent, knowing the play of others, determines his optimal strategy via a simple calculation. In the absence of this knowledge, any desire of a player to behave optimally requires a guess about the way in which others will play. For proposers in the UG such a guess is not only harder to come up with, but is also more important for their payoff than it is in the YNG. Proposers in the UG face much *strategic risk*, which is risk stemming from having a partner whose decisions affect their payoff. In the UG there is true risk from not knowing how the responder will behave: for each choice of the responder, a different payoff-maximizing behavior is prescribed for the proposer; in the YNG there is no such doubt (proposers have a weakly dominant strategy).[2]

This key difference between the UG and the YNG has been the target of several experimental studies (Güth and Kocher [2014] provides an excellent recent survey; Gehrig et al. [2007] and Güth and Kirchkamp [2012] are experimental studies of the YNG). These studies find that behavior in the YNG differs from behavior in the UG: proposers make smaller offers and responders are less willing to reject. Such a result suggests that strategic risk plays an important role in explaining the fair behavior observed in the UG. This behavioral difference does not answer our question, which is whether agents faced with a

[1] We use the word 'proposal' to refer to a division of the 100ECU, e.g., [90, 10], giving 90 to the proposer and 10 to the responder. The word 'offer' is used to refer to the responder's share that is specified by the proposer's proposal. In the example, the offer is of 10ECU.

[2] In the YNG there is strategic risk since the responder may reject all offers. Thus, a proposal may yield the payoff prescribed by this proposal or it may yield a payoff of zero (risk). However, since this risk affects all proposals equally, it plays no role in the determination of the proposer's optimal behavior.

one-shot game engage in introspection. In order to tackle our question we add the measurement of *response time* (RT). If time elapsed before making a decision (RT) is a good proxy for reflection or thought invested in making that decision, then the hypothesis that agents engage in introspection implies that the UG will generate larger RTs than the YNG. The use of two games where the role of one of the players (the proposer) is held constant across games in everything except the role of strategic risk - triggering the need to have or *form* beliefs about the behavior of other players - is ideal for the study of our question. Our first hypothesis, asking whether players in one shot games engage in thought about the behavior of others thus simply boils down to testing whether RT is larger for proposers in the UG than in the YNG. We find that this is indeed the case.

Of course, it does not suffice to know that such introspection occurs in order to know how to analyze one-shot games. It is also important to understand what this introspection is. We will not be able to answer such a question, but we will be able to say what behaviors are prevalent with more reflection (longer RT) for each game and across games. Our second hypothesis postulates that both games have instinctive, focal behaviors, that are predominant among subjects with short RT, while subjects with long RT move away from such behaviors, producing a higher dispersion of choices. We find that focal behaviors indeed prevail among subjects with short RT in both games, but the move away from such behavior by subjects with long RT, goes in directions we find surprising. We find that, when generous offers by proposers are not profit enhancing and not sustainable in equilibrium (YNG), they are more frequent among proposers with large RT. On the other hand, in the UG large RT is associated to a higher frequency of less generous offers. However, in the UG the [50, 50] proposal that is modal among proposers with short RT, as well as the [60, 40] proposal that is modal among proposers with large RT, are both justifiable on the basis of selfish, payoff-related reasons. Our findings thus disagree with literature asserting that more thought leads to the disappearance of (instinctive) selfless behaviors (see, e.g., Knoch and Fehr [2007], Rand et al. [2012]). They also disagree with the finding that increased RT amounts to additional levels of reasoning (Arad and Rubinstein [2012]) as considered either in levels-of-reasoning models or in cognitive hierarchies models (Hauk and Nagel [2001], Camerer et al. [2004]). The finding that proposers in the UG are generous but this generosity is profitable given the behavior of responders, is not new (Güth and Kocher [2014]). Our finding that larger RT coincides with larger expected and realized payoffs of proposers given actual behavior of responders, is akin to the finding of Avrahami et al. [2013], with experience taking the place of RT.

We present the experimental design, with details about the theoretical properties of the considered games, the measurement of RT, and the hypotheses we test, in section 2. Our results are presented in section 3. We discuss the findings and the setup and we conclude in section 4.

2 Experimental Design

We report data from an experiment run with 378 subjects across 12 experimental sessions. All subjects were undergraduate students (from different disciplines) at Jena University, had no training in Game Theory and had no previous experience with bargaining experiments in the lab. Students were recruited using ORSEE 2.0 (Greiner [2004]). The experiment was programmed and conducted with the software z-Tree (Fischbacher [2007]) at the computer laboratory of the Max Planck Institute of Economics (Jena, Germany). Subjects received written instructions, which were also read aloud by a research assistant to ensure everyone understood them (full experimental instructions can be found in supplementary online material). No communication between subjects was permitted. Interaction between participants was anonymous but in sessions covering 2/3 of the analyzed data subjects were informed of the gender of their partner and in all analyzed sessions subjects knew whether their partner was *from Jena* or *not from Jena*. At the end of every experimental session, subjects were paid in cash according to their payoff in the game (plus a show-up fee of €2.5). Every session lasted about 30 minutes and the average earnings per subject were €7.5.

2.1 The Games

Approximately half of the subjects (192 subjects) participated in an *Ultimatum Game* experiment (UG) while the remaining 186 subjects participated in a *Yes-or-No Game* experiment (YNG). Subjects made their payoff-relevant choice only once. In both games each subject's payoff depended on his own choice as well as that of an anonymous and randomly-assigned partner. The two subjects involved in each game had different roles. The *proposer* proposed one out of nine possible divisions of 100 Experimental Currency Units (ECU, with an exchange rate of 10 ECU/€), while the *responder* decided on conditional acceptance or rejection of each possible proposal.[3] The two games differed only in the choices available to the responder. In the UG the responder had to choose whether to accept or reject each possible offer of the proposer one by one. In this way the responder was allowed to condition his acceptance of the proposer's offer on the exact value of this offer. Responders made their choices *in cold*, before knowing the actual offer of their partner.[4] If the responder made his accept/reject choices such that he rejected all offers where he received less than a certain share, and accepted all offers where he received that share or more, we called this share the *Minimal Acceptable Offer* (MAO).

In the YNG, the responder was asked to choose whether he would accept every offer

[3] We used neutral names, X (proposer) and Y (responder), for the two roles.

[4] This way of implementing the classical sequential UG is called the *strategy method* (Selten [1967]) because it implements the strategic representation of the game.

or reject every offer from the proposer. Also in this case responders made their choices in cold, without knowing the proposal of their partner.

Subjects' payoffs in both games depended on the choices of both partners: they got paid only if the offer was deemed acceptable by the responder's choice, in which case they each received a share as established by the proposer's proposal. We will use bracketed pairs to denote the available proposals: [90, 10], [80, 20], and so on. The first number in each pair denotes the share going to the proposer and the second number, the share going to the responder.

2.2 Response Times

For each subject and matched pair of subjects we record strategy choices, payoffs, and *response time* (RT). We define response time to be the time elapsed between the moment in which a subject is presented with the problem and the moment when a choice is made. Subjects were not pressed to make a fast choice but they were allowed to leave the lab only after all subjects made a choice.[5] We will compare subjects' response times across games. Since response times depend on all elements of a decision, including complexity of the interface that presents the problem and physical motions needed to express a choice, it is important that the presentation of the problem be as similar as possible across the two games we compare. We make a big effort for this to be the case for the *proposers* in our experiment. For the responders, not having the same strategy set across games, the interface for decision is less comparable. We hence will give only marginal attention to RT comparisons across games for responders.

2.3 Experimental Hypotheses and Games

We will focus our analysis on two hypotheses, which we test via RT values and their interaction with proposer choices in both games.

H(1): Introspection. Proposers in the UG have longer response times than proposers in the YNG.

The above stems from the hypothesis that players in one-shot games engage in introspection. With this we mean that in one-shot games, players, recognizing that the behavior of other players affects their own payoff, will dedicate thought to forming beliefs about such behavior. To deduce hypothesis H(1) from the previous statement, one must assume

[5]The measurement of response times (RT) has been used in psychology to study mental structures since the mid 19th century. It is the time elapsed between a visual or auditory stimulus and the response, choice, or decision that the stimulus calls for (see, e.g., Luce [1991], Jensen [2006]). Experiments in psychology are typically simple in the number and type of choices they ask subjects to make and measurement is very precise (up to the millisecond). Our use of RT is very different. Since measurement is less precise, we limit our use of RT to comparisons across situations.

that more thought, on average, implies longer RT. One must also notice that proposers across the two games have identical strategy sets and choice interfaces. Hence, any difference across games in the amount of thought dedicated by proposers to choice can only be interpreted as thought dedicated to responders' choices.

There are two ways in which the choices available to responders in the UG are worthy of more thought by *proposers* in that game than in the YNG. First, the sheer numbers: in the YNG responders have only two possible strategies while in the UG they have nine monotonic strategies.[6] Second, *strategic risk*: in the YNG the proposer has a weakly dominant strategy, which means there is one option that delivers a (weakly) better material outcome regardless of the choice of the responder. In the YNG strategic risk is low for the proposer since the choice of the responder does not affect the proposer's understanding of which of his/her available strategies is individually optimal.

In the UG strategic risk for the proposer is high since, to start with, proposers in the UG have no dominant strategy: for each monotonic strategy of the responder a different strategy of the proposer becomes optimal. But even more, the UG is a game with many equilibria and each of the nine monotonic strategies available to responders can be sustained as part of a Nash equilibrium of the game. Summarizing, in the UG responders' choices affect what choice is optimal for proposers (no dominant strategy for proposer) and it is hard to predict what choice a responder will ultimately make (all responder's MAOs are sustainable in Nash equilibrium).

Our second hypothesis is concerned with within game differences in behavior due to differences in RT. This relates to much of the work in RT (see Rubinstein [2007], Piovesan and Wengstrom [2009], and Rand et al. [2012], for a few important examples). Prior work leads us to speculate that pro-social behavior will be most immediate in situations where this behavior is sustainable in equilibrium, like in the UG. In the YNG, where there is a unique Nash equilibrium, we expect this equilibrium to be focal behavior and, hence, most immediate.

H(2): Within-Game Behavior and *RT*. In both games subjects with small RT have more concentrated behavior than subjects with large RT.

In particular, in the UG, behavior of proposers with small RT is modal around the egalitarian proposal [50, 50], while in the YNG, behavior of proposers with small RT is modal around the selfish proposal [90, 10] (unique equilibrium).

The reason to consider [50, 50] a focal equilibrium is evidence from experiments with the UG where this proposal is found to be modal. Moreover, proposers who believe that responders will reject unfair offers, will find that an offer of 50 is a safe bet, minimizing

[6] The total number of strategies available to responders in the UG is 2^9. Strategies for which a MAO exists – monotonic strategies – are only nine: one for each possible proposal by the proposer. It is reasonable and empirically justified to believe that responders will mainly use such monotonic strategies.

		Subject Group				
		Female	*Eq. gender*	From Jena	*Eq. origin*	All
Mean offer	UG	42.3 (39.6)	40.6 (42.0)	40.0 (41.0)	41.5 (38.3)	40.9
	YNG	22.2 (26.4)	26.4 (23.0)	21.4 (24.6)	25.6 (21.1)	24.3
Mean MAO	UG	22.8 (22.9)	22.2 (22.5)	26.7 (22.3)	23.3 (21.0)	22.8
Mean RT **proposers**	UG	32.7 (29.1)	29.5 (31.4)	29.7 (31.0)	29.1 (37.5)	30.8
	YNG	24.6 (26.2)	24.5 (22.4)	38.4 (24.3)	23.8 (29.4)	25.4
Mean RT **responders**	UG	38.5 (41.9)	41.2 (39.0)	45.1 (39.5)	40.6 (38.7)	40.2
	YNG	15.8 (14.6)	14.7 (14.6)	15.0 (15.3)	15.0 (15.8)	15.2

Table 1: Summary statistics of behavior and RT given the known demographic data of participants. Alternative in parenthesis: Female (Male), Eq. gender (Not eq. gender), From Jena (Not from Jena), Eq. origin (Not eq. origin).

strategic risk while maintaining a reasonable payoff for the proposer.

There are several reasons why subjects with larger RT are expected to display less concentrated behavior, and we will only be able to somewhat uncover these reasons. We name a few. First, subjects who understand what is selfishly optimal may nonetheless feel morally attracted to make different offers. This moral dilemma may be a cause of delay and the final resolution of the dilemma may be a generous offer or not. Second, subjects who do not understand well the game may be slower and their behavior is likely to be noisy. Finally, subjects who understand the trade off between payoff and strategic risk may try to outguess the behavior of responders or simply gather the courage to make a proposal different from the egalitarian proposal in the UG. Here we will provide evidence of dispersion in behavior for large RT, and we will give suggestive evidence for the first and last of the spelled out reasons.

3 Results

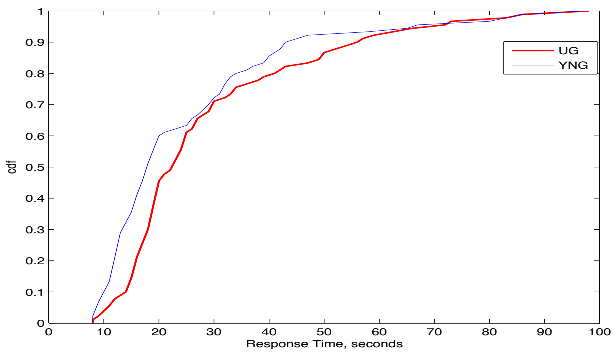
Table 1 contains descriptive statistics for our data divided by gender and origin. In all our sessions subjects were told whether their partner was from Jena or not and in 2/3 of all considered data subjects were told the gender of their partner. Approximately 50% of proposers and responders in both games were female and, hence, approximately 50% of all matches were among subjects of equal gender (in those sessions where subjects were told the gender of their partner). 12% of both proposers and responders were from Jena in the UG, 8% of proposers and 23% of responders were from Jena in the YNG. Approximately 75% of all matches were among subjects who were both from Jena or both not from Jena. As is suggested by the data in table 1, gender, origin, and homogeneity of the match had no qualitative effect on RT and behavior. Wherever relevant we further substantiate this claim with the appropriate statistical analyses.

		RT Quartile			
		(Fast)	$\longrightarrow$		*(Slow)*
		$Q1$	$Q2$	$Q3$	$Q4$
KL distance from uniform	UG	0.56	0.42	0.40	0.31
	YNG	0.62	0.43	0.24	0.04
Pearson's $\chi^2 - H_0$: Uniform	UG	26.8**	16.8**	13.4**	13.3*
	YNG	31.6**	18.4**	11.6*	2.0

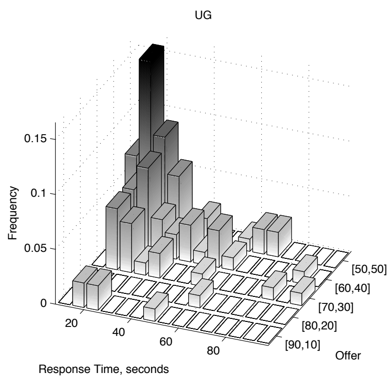
Table 2: Upper panel: Kullback-Leibler measure of divergence from a uniform distribution over all offers. High values indicate concentration, low values indicate dispersion. Lower panel: Pearson's χ^2 test of equality to a uniform distribution over all offers. (**) Difference from a uniform distribution significant at the 1% level. (*) Difference from a uniform distribution significant at the 5% level.

We cannot reject hypothesis **H(1)**. Overall *RT*s are larger in the UG than in the YNG for both proposers and responders. The mean proposer *RT* in the UG is 30.8 seconds, while it is 25.4 seconds in the YNG, significantly smaller than the UG (the t-test of equal means against the alternative hypothesis that the YNG has smaller *RT*, has p-value 0.0415). Moreover, the Kolmogorov-Smirnov test of first order stochastic dominance of the distribution of *RT* for the UG over that for the YNG has a p-value of 0.015. This first order stochastic dominance is graphically displayed in figure 1a. If we restrict attention to proposal [50, 50], which is focal in the UG but not in the YNG, we still find that YNG response times are shorter, albeit not significantly so (mean of 20.24 seconds in the YNG and 26.27 in the UG, one-sided p-value for difference of means equal to 0.12). To control for gender, origin, and matching effects (by gender and origin), we ran regressions of *RT* on a game dummy variable (simple t-test), but included controls capturing gender, origin, and particular matching among participants, like male-with-male, etc. (t-test with controls). The difference in *RT* between games remains significant (one-sided t-test p-value less than 0.04 in all regressions, just like in the t-test without controls – see section 5 in the supplementary online material for regression reports).

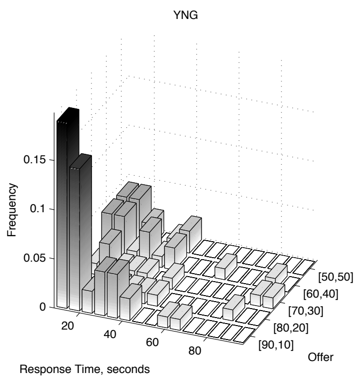
We now turn to hypothesis **H(2)**, which states that in both games there are focal proposals made by proposers with small response times, while proposers with large response times make a bigger variety of proposals. We therefore now focus on behavior (proposals) of proposers in our two considered games. In the UG, overall, the modal proposal is [50, 50], while it is [90, 10] in the YNG. The distributions of proposals across games are statistically different (Pearson's χ^2 test with p-value less than 0.001). There is a negative rank correlation between the popularity of a proposal and the *RT* of subjects making such a proposal ($\rho = -0.179$, significant at the 10% level): frequent choices are also the choices made by proposers with small response times, which supports the existence of *instinctive* choices. The *instinctive* choice for the UG is to propose [50, 50], while for the YNG it is



(a) Empirical cumulative distribution function (CDF) of RT for the UG and the YNG.



(b) UG: Joint histogram of RT and Offer.



(c) YNG: Joint histogram of RT and Offer.

Figure 1: Distributional data of RT and Offer for proposers in both games. RT in the UG *first-order stochastically dominates* RT in the YNG (figure 1a). Figures 1b and 1c show how offers differ across games and depending on proposer's RT.

		Offers			RT	
		Mean	*Median*	*Mode*	*Mean*	*Coeff. of variation*
UG Proposers	$Q1$	43.0	50	50	14.1	0.19
	$Q4$	39.1	40	40	61.6	0.43
YNG Proposers	$Q1$	21.7	10	10	10.9	0.15
	$Q4$	27.0	30	10	50.3	0.40
UG Responders	$Q1$	16.3	10	10	23.2	0.15
	$Q4$	30.4	30	40	67.3	0.28

Table 3: Comparison of offers to the responders and RT between RT quartiles $Q1$ and $Q4$.

to propose [90, 10]. We find a significant percentage (approx. 52%) of offers larger than 10 in the YNG. In the extant literature this percentage is often even larger and is attributed to simple inequality aversion (Gehrig et al. [2007]). Slower choices are less concentrated, as we show in table 2.[7]

Figures 1b and 1c show the joint histogram of behavior (proposer offers) and RT, hinting at the fact that there are significant behavioral differences between differently fast proposers in each of the considered games and that non-focal proposals are the most frequent among proposers with large RT. Table 2 shows data divided into four speed quartiles (quartile 1, $Q1$, contains the proposers with smallest RT, and quartile 4, $Q4$, contains the proposers with largest RT), and table 3 shows data for the first and last of these quartiles.[8] Table 2 contains two indicators of the dispersion of choice. Clearly, dispersion increases with higher quartiles in both considered games.

We already saw that in both games there are instinctive choices that prevail among proposers with small values of RT, and that these choices are different across games. In what remains we analyze what are the choices made by proposers with large values of RT.

Table 3 compares behavior and payoffs of subjects in $Q1$ and $Q4$. In the YNG, the distribution of proposals in $Q4$ is not distinguishable from a uniform (uniform distribution over the 5 proposals giving the proposer 50 or more). Median proposals are [90, 10] in $Q1$ and $Q2$, [80, 20] in $Q3$, and [70, 30] in $Q4$. There is a strong correlation between RT and pro-social proposals that are neither equilibrium nor individually rational in terms of

[7] Proposals across games are statistically different after controlling for gender, origin, matching on gender and origin, and cross effects of these control variables. Given the nature of the dependent variable *offer*, we ran *interval regressions* of offer as a function of a game dummy variable and the control variables. The t statistic for game effect has a p-value less than 0.001 regardless of the controls considered and controls are never significant. An interval regression is a maximum likelihood estimation where the likelihood of each data point is the density evaluated on an interval instead of at a point. Full regression reports in the supplementary online material.

[8] The categorization is approximate whenever 25% of the considered subset of the data is not an integer. An alternative categorization, used both in Rubinstein [2008] and Piovesan and Wengstrom [2009], is to split the data into four unequally-sized groups: the *Very Fast* 10% with smallest values of RT, the *Very Slow* 10% with largest values of RT and two groups containing 40% of the data each, of *Fast* and *Slow* subjects. Our qualitative results are maintained with this alternative categorization.

payoff. We cannot disentangle whether the high frequency of off-equilibrium, pro-social proposals among proposers with high RT in the YNG is due to a correlation between RT and mis-understanding of the game or due to time required to solve a (moral) dilemma among pro-social proposers. The fact that no equivalent statistical equality to the uniform distribution arises in the UG, which is a more complex game, plays in favor of the latter interpretation.

In the UG, the distribution of proposals in $Q4$ has full support and the modal proposal is significantly less accentuated than in $Q1$.[9] The median proposal is [50, 50] in $Q1$ and [60, 40] in all other quartiles, while the mode is [50, 50] in $Q1$ and is [60, 40] in $Q4$. The latter is particularly important in view of the behavior of responders in the UG.

Given the empirical distribution of MAO, proposal [60, 40] has the largest expected payoff (ECU 58.75) and a lower variance than all lower proposals (it has larger variance than proposal [50, 50], which yields a sure payoff of 50 in our sample). Given the distribution of MAO that we observe, we find that on average, proposers in $Q4$ made choices with the highest expected payoff (table 3). In the UG, unlike in the YNG, larger RT correlates with more individually rational (in terms of payoffs) behavior. Choices of proposers with large RT are strategically riskier but yield a higher payoff in expectation, given responders' behavior.

Until this point we had disregarded responder data (except for the calculation of proposal expected payoff in the previous paragraph), because of the difficulty of comparing responder RT data across games.[10] On the other hand, the comparison of responder behavior (MAO) across RT quartiles is not only possible but also very useful. There are significant behavioral differences across quartiles,[11] with a positive correlation between RT and MAO. Like the proposers in the UG, the responders with large RT are more prone to make risky choices, but unlike these proposers, such choices are not in and of themselves payoff enhancing. Like the proposers in the YNG, responders with large RT in the UG are likely to have social motivations: a preference for fairness, spite, or the social responsibility to enforce a responder-favorable equilibrium.

[9] If the RT categories of Rubinstein [2007] and Piovesan and Wengstrom [2009] are used, the behavioral difference between the *Very Fast* and the *Very Slow* is more accentuated than that between our $Q1$ and $Q4$. While the *Very Slow* proposers make every possible proposal, the *Very Fast* make only proposals [50, 50] and [60, 40].

[10] All the comparisons of responder RT across games go in the expected direction. Responder RT in the UG is larger on average and more dispersed than responder RT in the YNG. Responders in the UG also have larger response times, on average, than proposers.

[11] Pearson's χ^2 test with p-value$= 0.018$, regressions with or without control variables yield coefficients $Q3$ and $Q4$ dummies that are significant at 1%, while $Q2$ is never significantly different from $Q1$ (omitted, base).

4 Discussion

In our work we find that, on average, proposers have larger RT in the game that is strategically more complex (UG), even though the strategy space and interface of the two considered games are identical. Across games (UG and YNG) and roles (proposer and responder), larger RT is associated with behaviors where there is a risk or a certainty of loosing money with respect to the *instinctive* choice (choice made by the majority of subjects, and most so by subjects with a small RT). However, the motivation behind choosing to bear such a (potential) loss of money differs significantly across games and roles. For proposers in the UG, it leads to larger payoffs in expectation, due to a better reply to responder behavior; for proposers in the YNG, it leads to a sure loss of money from generous, pro-social behavior. For responders in the UG, it indirectly leads to larger payoffs: it is the behavior of responders with large RT that sustains proposers' generous behavior, by inducing a risk of rejection for low offers.

Since our paper is concerned with behavior in one-shot games, it is obvious to ask how our data relate to models of levels of reasoning and cognitive hierarchies (Nagel [1995], and Camerer et al. [2004], respectively). We do not find that behavior of subjects with larger values of RT corresponds to higher levels of reasoning.[12] Arad and Rubinstein [2012] find the opposite. This is likely due to a few important differences between our work and theirs: the games we consider are more social and simpler than the Colonel Blotto game analyzed in Arad and Rubinstein [2012], and the levels of reasoning model used in their work is modified to include a stage where the problem is broken down in separate dimensions.

Arad and Rubinstein [2012] make an effort to strip off of social content the situation faced by their experimental subjects. Our work and findings are closer to works dealing with the cross effects of reflection (RT in our case) and social concerns on behavior (e.g., Knoch and Fehr [2007], Piovesan and Wengstrom [2009], Rand et al. [2012], Fischbacher et al. [2013]). Unlike in Rand et al. [2012], we do not find that reflection leads to a depletion of intuitive cooperative attitudes. On the contrary, in our experiment pro-social behavior with negative payoff effects appear only among proposers in the YNG and responders in the UG that have large response times. Rand et al. [2012] remark, however, that the effect of reflection on cooperative attitudes is modulated by their subjects' experience in life and with experiments: instinctive cooperation in their setting may be driven by subjects who associate the game in the experiment with games they play in life where they and their

[12] When the UG is analyzed with a levels of reasoning model, whenever a proposer of level j proposes [50, 50], a responder of level $j + 1$ is indifferent between all values of MAO between 10 and 50. If we assume that a responder who is indifferent randomizes uniformly over all strategies, we get oscillation: proposers of odd level propose [50, 50], proposers of even level propose [90, 10]; responders of odd level set a MAO of 10, responders of even level randomize equally over all values of MAO. A cognitive hierarchies model, where players of level j best respond to a population containing players with all levels lower than j, predicts convergence to proposal [90, 10] by proposers of high level, and to a MAO of 10 by responders of high level.

co-players cooperate. Among subjects who are experienced with cooperation experiments, they find that *RT* may actually be enhancing of cooperation (this finding is not significant). In a study using temporary impairment of brain functions via magnetic stimulation, Knoch and Fehr [2007] find that UG responders impaired in areas associated to self control are less likely to reject low offers. This relates to our finding that small response times associate to small values of MAO. Rejecting unfair offers requires more self control (Knoch and Fehr [2007]) or longer *RT* (our study). In a very innovative study, Fischbacher et al. [2013] address *procedural heterogeneity* among subjects. They find that different types of subjects (as identified from behavior in several ultimatum games) display different behavioral effects of RT. Our work suggests that a similar heterogeneity may arise across games.

Finally, Piovesan and Wengstrom [2009] study dictator games, where strategic considerations are completely removed while maintaining the social nature of the situation. Just like in our YNG, they find that larger *RT* is associated with more pro-social behavior. Their across-games analysis shows that *RT* is larger if the *moral dilemma* faced by dictators is more complex (there is no clear *fair* option). A comparison of *RT* between the UG, the YNG, and the dictator game may be illuminating. In the dictator game there is no strategic risk for the proposer since the proposer's payoff depends on his/her choice only. Behaviorally, on average, offers in the dictator game are lowest, followed by YNG offers and finally UG offers (Güth and Kocher [2014]). One may expect a similar ranking of the games in terms of proposer RT. However, the dictator game also poses a very different social and moral dilemma than the UG and YNG, since the proposer (dictator) bears full responsibility for the payoff of the responder. When we take into account the correlation we find between pro-social behavior and *RT* in the YNG, and the changed social role of the proposer in the dictator game, the expectation of *RT* for the dictator game becomes less clear. Therefore, adding this third game for comparison may deliver surprising and informative results.

Since we are concerned with one-shot games, we analyze data on one-shot experimental games. However, there is much to be learned from the evolution of the relation between *RT* and behavior when experience is accumulated. Only Piovesan and Wengstrom [2009] brush over the analysis of such data. Whatever introspection is used by subjects to choose behavior in novel strategic situations should become less relevant as they accumulate experience. Hence, it is of interest to understand what behavioral effects of *RT* disappear with experience. In our case, will *RT* still correlate with pro-social behavior? Will it correlate with the taking of (strategic) risks? We leave such questions to future work.

References

Ayala Arad and Ariel Rubinstein. Multi-dimensional iterative reasoning in action: The case of the colonel blotto game. *Journal of Economic Behavior and Organization*, 84(2): 571–585, 2012.

Judith Avrahami, Werner Güth, Ralph Hertwig, Yaakov Kareev, and Hironori Otsubo. Learning (not) to yield: An experimental study of evolving ultimatum game behavior. *The Journal of Socio-Economics*, 47:47–54, 2013.

Colin Camerer, Teck-Hua Ho, and Juin-Kuan Chong. A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898, 2004.

Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10:171–178, 2007.

Urs Fischbacher, Ralph Hertwig, and Adrian Bruhin. How to model heterogeneity in costly punishment: Insights from responders' response times. *Journal of Behavioral Decision Making*, 26(5):462 – 476, 2013.

Thomas Gehrig, Werner Guth, Vittoria Levati, Rene Levinsky, Axel Ockenfels, Tobias Uske, and Torsten Weiland. Buying a pig in a poke: An experimental study of unconditional veto power. *Journal of Economic Psychology*, 28:692–703, 2007.

Benjamin Greiner. An online recruitment system for economic experiments. In K. Kremer and V. Macho, editors, *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*. Ges. fur Wiss., 2004.

Werner Güth and Oliver Kirchkamp. Will you accept without knowing what? the yes-no game in the newspaper and in the lab. *Experimental Economics*, 15:656–666, 2012.

Werner Güth and Martin Kocher. More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior and Organization*, 108:396–409, 2014.

Esther Hauk and Rosemarie Nagel. Choice of partners in multiple two-person prisoner's dilemma games. *Journal of Conflict Resolution*, 45(6):770–793, 2001.

Arthur Jensen. *Clocking the Mind: Mental Chronometry and Individual Differences*. Elsevier, The Netherlands, 2006.

Daria Knoch and Ernst Fehr. Resisting the power of temptations. the right prefrontal cortex and self-control. *Annals of the New York Academy of Science*, 1104:123–134, 2007.

Robert Duncan Luce. *Response Times: Their Role in Inferring Elementary Mental Organization.* Oxford University Press, 1991.

Rosemarie Nagel. Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326, 1995.

Marco Piovesan and Erik Wengstrom. Fast or fair? a study of response times. *Economics Letters*, 105:193–196, 2009.

David Rand, Joshua Greene, and Martin Nowak. Spontaneous giving and calculated greed. *Nature*, 489:427–430, 2012.

Ariel Rubinstein. Instinctive and cognitive reasoning: A study of response times. *The Economic Journal*, 117:1243–1259, 2007.

Ariel Rubinstein. Comments on neuroeconomics. *Economics and Philosophy*, 24(485–494), 2008.

Reinhard Selten. Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperiments. In Heinz Sauermann, editor, *In Beiträge zur experimentellen Wirtschaftsforschung*, volume I, pages 136–168. J.C. Mohr (Siebeck), 1967.

Strategic Risk and Response Time Across Games
Supplementary Online Material

Pablo Brañas-Garza[*] Debrah Meloso[†] Luis Miller[‡]

June 15, 2015

1 Experimental Instructions

Welcome and thank you very much for participating in this experiment. Please read the instructions carefully. If you have any questions or concerns, please raise your hand. It is strictly forbidden to communicate with other participants during the experiment. It is very important that you follow this rule. Otherwise we must exclude you from the experiment and from all payments. Should you have any question, please raise your hand and we will answer it individually.

During the experiment, we use ECU (Experimental Currency Unit) instead of euro. At the end of the experiment, the ECU you have earned, will be converted to euro (10 ECU = 1€) and the obtained amount will be paid to you in cash.

In this experiment, two participants will interact with each other just once. Each of the two members of a pair will be randomly assigned one of two roles: X or Y. In the top right corner of the computer screen, you can read which role (either X or Y) has been assigned to you and to your partner.

Each pair can share 100 ECU. X has the right to propose the distribution of the 100 ECU. In particular, X chooses the distribution (x, y) meaning that X wants to keep x ECU for him/herself, and to give y ECU to Y. More specifically, X can choose any of the following 9 distributions:

x	10	20	30	40	50	60	70	80	90
y	90	80	70	60	50	40	30	20	10

[*]Middlesex University Business School, Hendon Campus, The Burroughs, London NW4 4BT, UK
[†]ESC-Rennes School of Business **Corresponding author**. debrah.meloso@esc-rennes.com
[‡]BRIDGE & DFAE II, Universidad del País Vasco (UPV/EHU), Av. Lehendakari Aguirre 83, 48015 Bilbao, Spain

Ultimatum Game Only
Y must decide for each possible distribution of the 100 ECU, if he or she accepts or rejects it. Thus, Y will face the following table:

x	10	20	30	40	50	60	70	80	90
y	90	80	70	60	50	40	30	20	10
Accept									
Reject									

For each possible distribution, Y must specify if he or she accepts or rejects it by checking the corresponding box (thus Y is required to make 9 decisions). After X and Y have made their choices, their payoffs are determined as follows:

- If Y has accepted the actual proposal by X, then both get what X has proposed, i.e., X earns x and Y earns y.
- If Y has rejected the actual proposal, then both earn nothing, i.e., the 100 ECU are lost.

Yes-or-No Game Only
Without knowing which of the 9 possible proposals X has chosen, Y must accept or reject it.
After X and Y have made their choices, their payoffs are determined as follows:

- If Y has accepted, then both get what X has proposed, i.e., X earns x and Y earns y.
- If Y has rejected, then both earn nothing i.e., the 100 ECU are lost. It must be emphasized that Y does not know the actual distribution (x, y) proposed by X when deciding whether to accept or reject it.

At the end of the experiment, the actual payoff will be paid out in cash, together with the show-up fee of €2.50 for having shown up on time.

2 Game Presentations

Figures 1 to 3 display the way in which the game was presented to subjects in our experiment. Notice that the screen for proposers in the UG and the YNG was exactly the *same*.

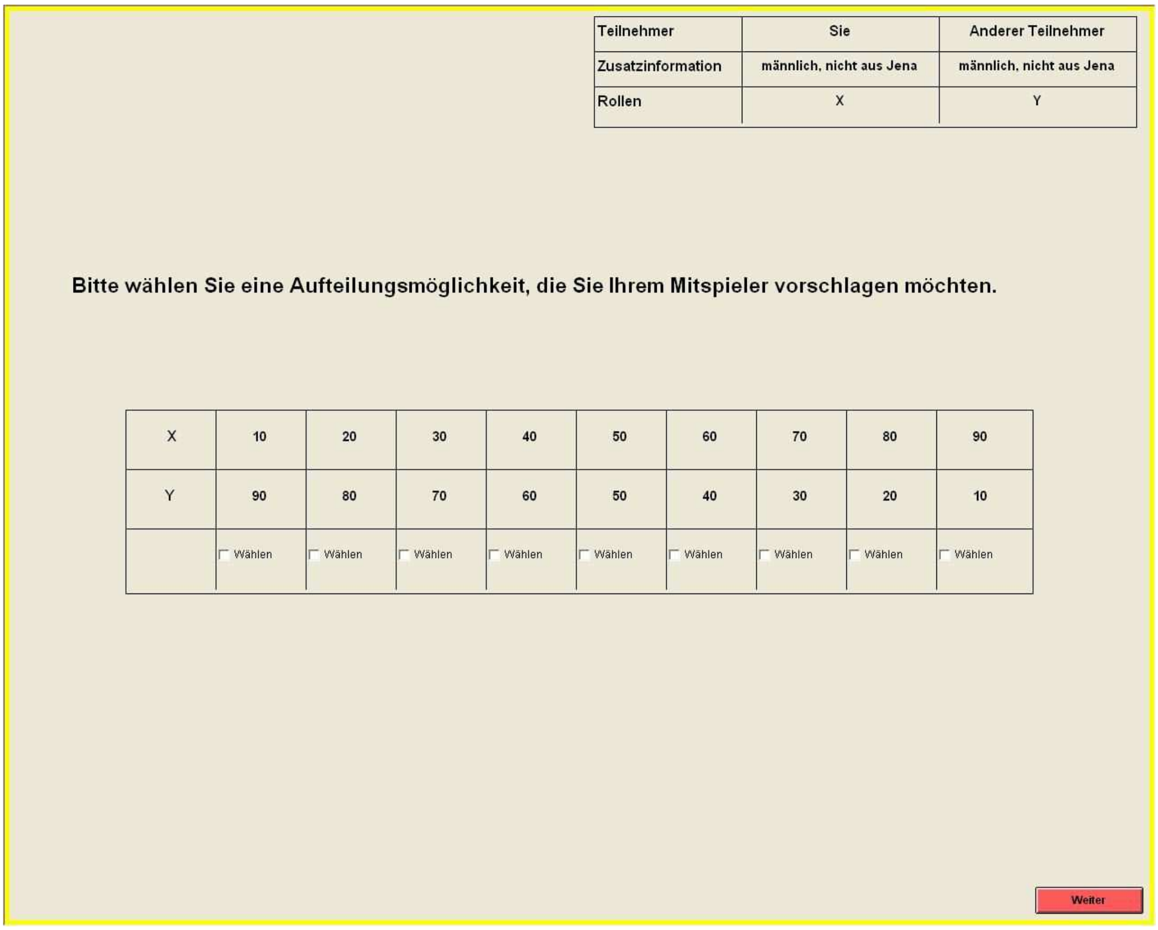


Figure 1: Screen where proposers in either the UG or the YNG were presented their available options and made their choice.

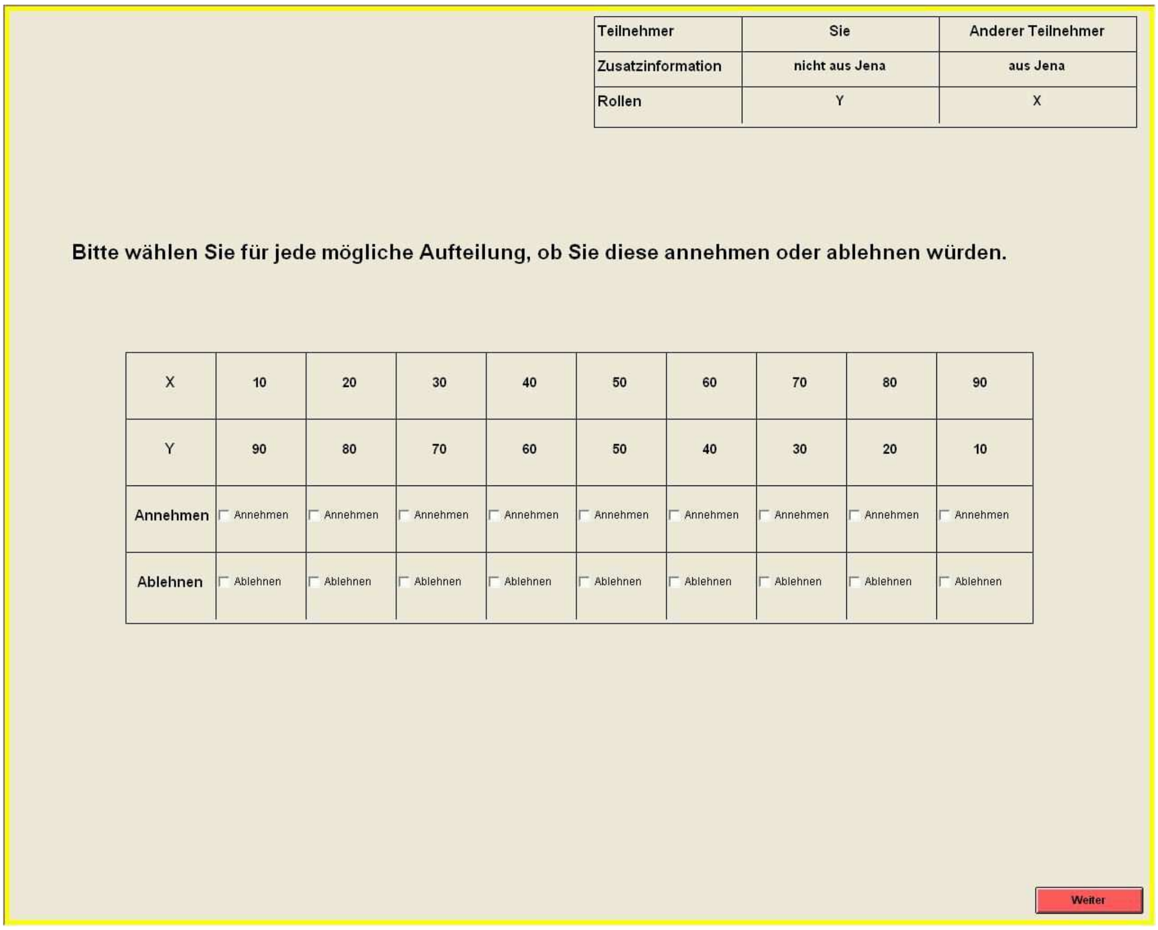


Figure 2: Screen where responders in the UG were presented their available options and made their choice.

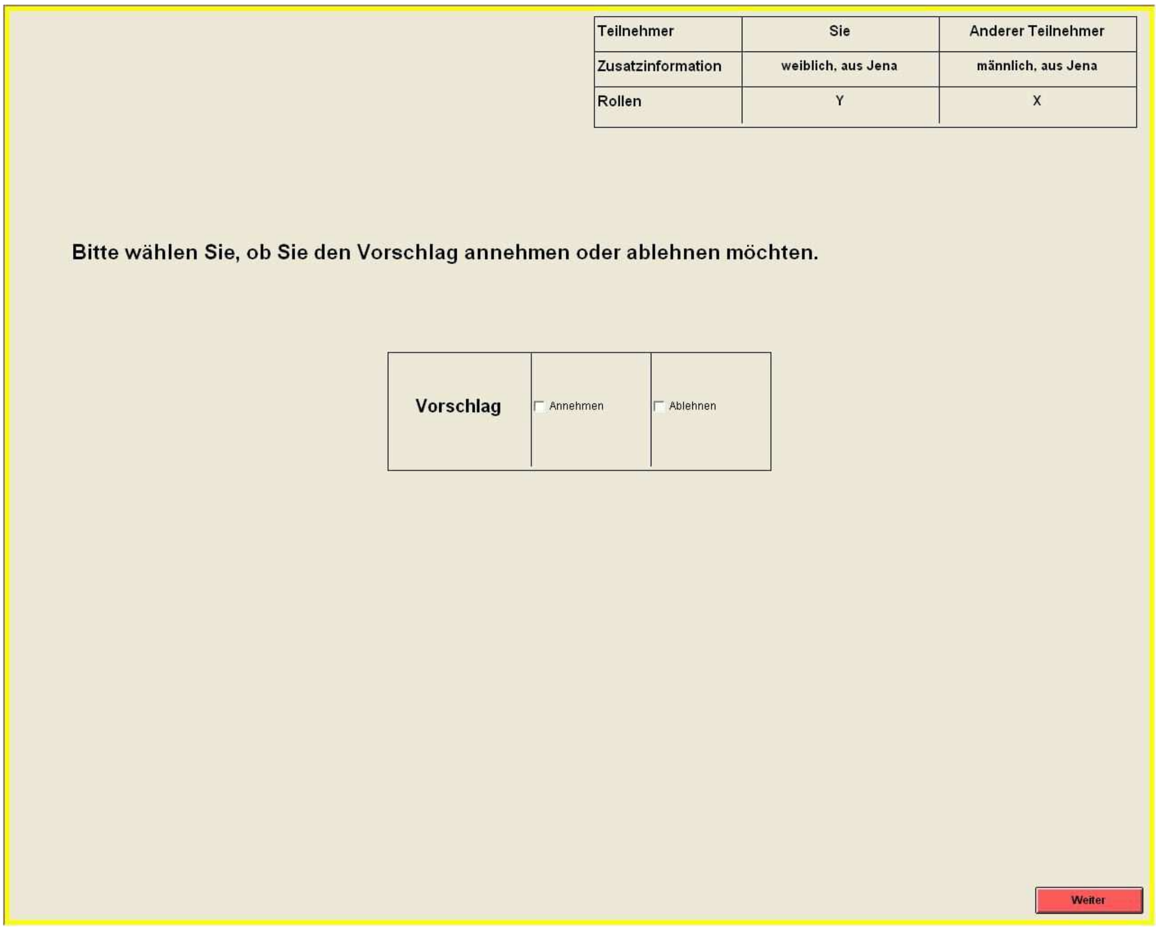


Figure 3: Screen where responders in the YNG were presented their available options and made their choice.

3 Further Details of Collected Data

In the YNG we had a total of 186 participants, 94 in the role of responders and 92 in the role of proposers. The mismatch is due to an inconsequential technical issue that matched two responders to the same proposer in two separate occasions. Subjects involved in this situation were not aware of it and both responders in both occasions "accepted every offer", meaning that the decision for the payoff of their proposer was unambiguous. Not all YNG data were used: there were two proposers who made offers larger than 50, we excluded them from all data analysis.

In the UG we had a total of 192 participants, 96 in each role. Not all data were used. We have five proposers who make offers larger than 50, we exclude them from all statistical analysis. We also have four responders who do not have a minimal acceptable offer (MAO) since they submitted non-monotonic strategies. The four instances are: Subject 321, who rejects offers where he receives 90, 10, or 20. Subject 414 accepts the offer where he receives 10 and rejects all other offers. Subject 513 rejects offers where he receives more than 60. Finally, subject 1308 only accepts offers where he receives either 40, 50, or 60.

Tests repeated using all proposer data, without excluding offers larger than 50, for both the UG and the YNG, led to minor quantitative changes with no qualitative consequences.

4 Gender and Origin Effect on Proposals

In table 1 we report interval regressions of proposer offers on a game dummy and the controls of gender, gender match, origin (either from Jena or from elsewhere), origin match and cross effects. Gender is measured with a dummy variable taking value 1 if the proposer was female (variable $female$). Gender match is measured with a dummy variable taking value 1 if the proposer was matched with a responder of the same gender (variable *gmatch*). Origin is measured with a dummy variable taking value 1 if the proposer was **not** from Jena (variable $foreign$). Origin match is measured with a dummy variable taking value 1 if the proposer was matched with a responder of the same origin; i.e., both proposer and responder were from Jena or both proposer and responder were not from Jena (variable *omatch*). Finally, cross effects of both gender match and origin match were measured with a dummy variable called *gomatch*.

Interval regressions are maximum likelihood estimations where data points are matched to an entire interval and its probability instead of to a single point. Being maximum likelihood estimations, we do not give goodness of fit for each model using R^2, but using instead the χ^2 estimator for the likelihood ratio between the estimated model and a model with only a constant. In the models reported in table 1, the likelihood ratio test is significant, meaning that the models are useful beyond computing a simple mean. This is not the case

	Models		
Variables	Model 1	Model 2	Model 3
game	-16.523^{***}	-16.085^{***}	-16.019^{***}
	(1.996)	(2.019)	(2.020)
female	-0.718		-0.363
	(1.997)		(2.017)
gmatch	1.430		-0.360
	(2.102)		(4.249)
foreign		-1.128	-1.228
		(3.731)	(3.761)
omatch		4.296	3.475
		(2.670)	(3.152)
gomatch			2.373
			(4.885)
Constant	40.727^{***}	38.408^{***}	38.77^{***}
	(1.84)	(3.194)	(3.720)
Observations	181	181	181
Likelihood ratio χ^2	58.67^{***}	60.96^{***}	61.72^{***}

Table 1: Interval regressions of proposal as a function of game type (UG or YNG, UG used as base), and several control variables. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$.

if we drop the game dummy from regressions.

In table 2 we report interval regressions of proposal on the control variables only, with the data separated by games. If we do not separate by games (not reported here) the variance of the coefficients of the control variables is even larger and the qualitative result of no explanatory power holds through. As can be seen in table 2, none of the control variables has a significant effect on the value of proposals, and the models have no meaningful explanatory power beyond the computation of a simple mean of proposals (likelihood ratio test is never significant, not even at the 0.1 level). This is true for both the UG and the YNG.

5 Gender and Origin Effect on Response Tiimes

Table 3 contains t-tests of the difference in mean RT of proposers across games, without controls, as well as with different controls of gender and origin of subjects ($female$, *gmatch*, $foreign$, *omatch*, and *gomatch* as defined in section 4). The p-values for the treatment effect (variable YNG) are given for a one-sided test of the difference in means – the alternative hypothesis being that RT is larger in the UG than in the YNG.

	UG			YNG		
Variables	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
female	2.760		2.906	−4.150		−3.591
	(2.243)		(2.245)	(3.266)		(3.346)
gmatch	−0.407		−3.521	3.128		1.922
	(2.362)		(5.559)	(3.437)		(6.227)
foreign		−2.156	−1.573		0.454	−1.076
		(4.225)	(4.219)		(6.581)	(6.606)
omatch		4.305	3.321		4.372	3.536
		(3.586)	(3.955)		(3.889)	(4.738)
gomatch			3.502			2.101
			(6.127)			(7.477)
Constant	39.70***	39.282***	38.392***	25.333***	20.807***	23.456***
	(1.730)	(3.143)	(3.720)	(2.613)	(5.912)	(6.783)
Observations	91	91	91	90	90	90
Likelihood ratio χ^2	1.52	1.52	3.54	2.46	1.51	3.83

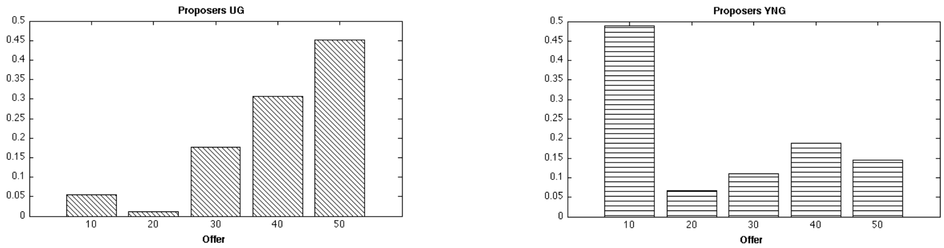
Table 2: Interval regressions of proposal as a function of several control variables. Data separated by game. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$.

6 Response Time and Behavior Figures

Figure 4a clearly indicates that the most frequent offer in the UG is [50, 50], and figure 5a indicates that this is overwhelmingly the modal offer among the *Very Fast* proposers. Figures 5a and 5b show histograms of proposer behavior (offers) divided into four categories according to RT. Each category contains approximately 25% of the data (slightly more or slightly less when the total data are not divisible by four). The *Very Fast* category contains behavior for the 25% of proposers that have the smallest RT, with growing RT up to category *Very Slow*.

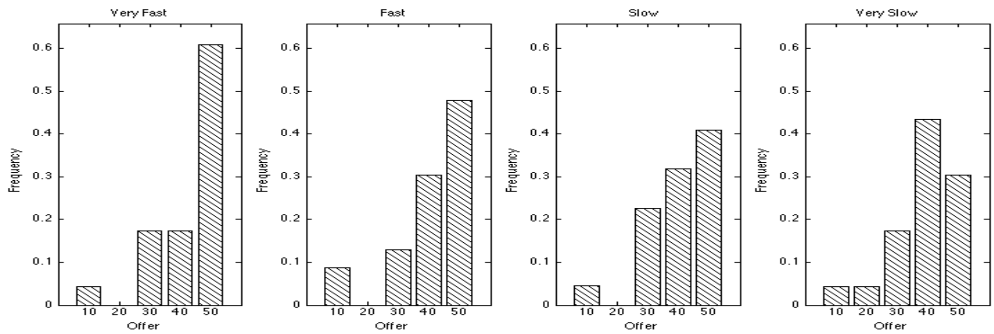
Proposer behavior in the YNG differs significantly from that in the UG, both overall and by speed category. The modal offer is [90, 10], and this is particularly accentuated among the fastest proposers (see figures 4b and 5b).

Figure 6 contains information on RT and the joint histograms of RT and behavior (MAO) for responders in the UG. Figure 7a shows histograms of MAO across all RT categories. The table shows the expected payoff of each offer the proposers can make, that is implied by the distribution of responder behavior. That is, given the MAO set by responders in our dataset, each offer of the proposers in the UG may be rejected or not. Given these rejection rates and the payoff that each proposal gives to the proposer him/herself, the proposals will have different expected payoffs that are summarized in the table in figure 7b (G_x denotes probability of acceptance of offer x).

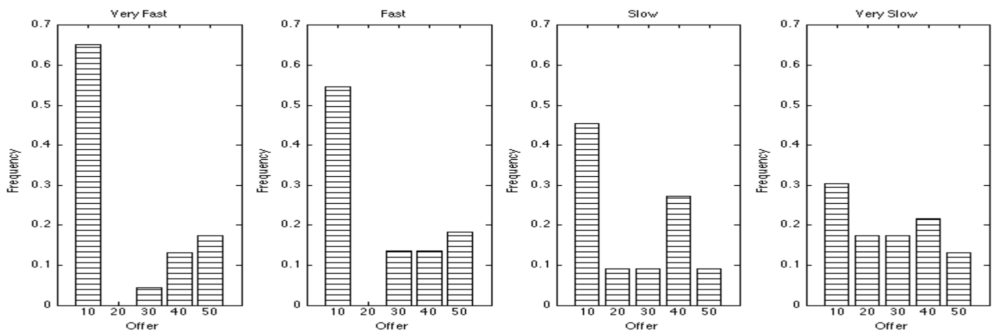


(a) Proposer offers in the UG. (b) Proposer offers in the YNG.

Figure 4: Histograms of offers $[100 - x, 100]$, where $x \in \{10, 20, 30, 40, 50\}$.

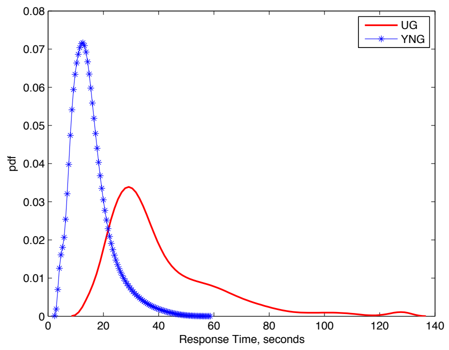


(a) Histograms of offers in each *RT* category for proposers in the UG.

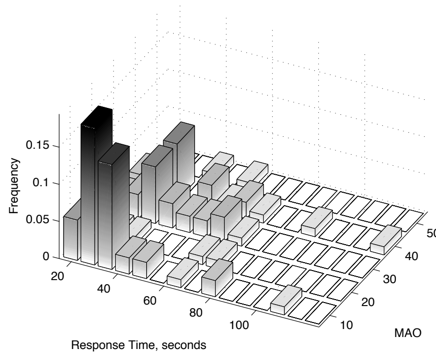


(b) Histograms of offers in each *RT* category for proposers in the YNG.

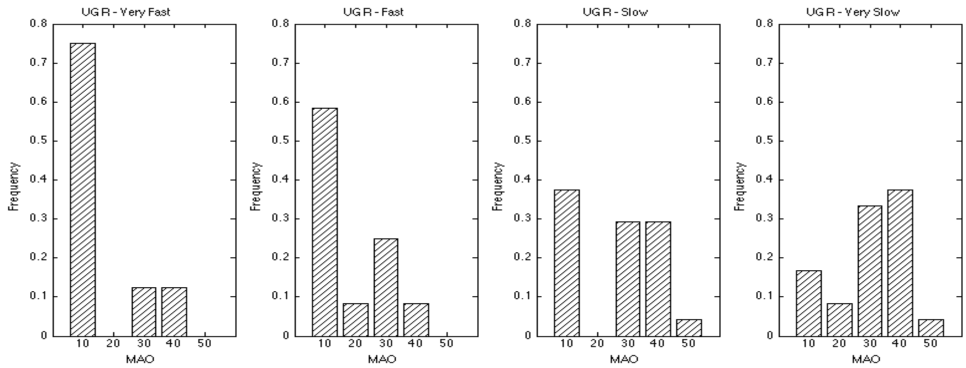
Figure 5: Four categories according to proposer *RT*: There are 9 subjects in *Very Fast* and in *Very Slow*, 36 subjects in *Slow*, and 36 in *Fast* in the YNG, and 37 in the UG.



(a) Empirical PDF of RT for responders in either game.

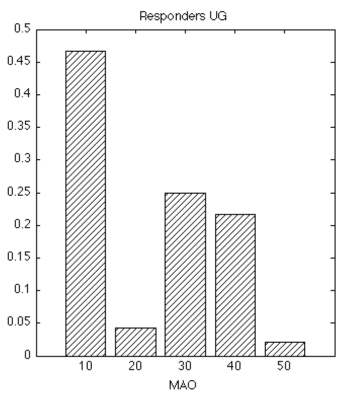


(b) Joint histogram of responder MAO and RT in the UG



(c) Histogram of MAO for responders in different RT categories.

Figure 6: Data on choices and RT for responders.



(a) Histogram of responder MAOs.

Proposer's offer of $[100 - x, x]$

$x \rightarrow$	10	20	30	40	50
G_x	*0.4688*	*0.5104*	*0.7604*	***0.9792***	*1*
P_x	42.19	40.83	53.23	**58.75**	50
SE_x	44.91	39.99	29.88	**8.57**	0

(b) Distribution of payoffs for each offer, given the distribution of MAOs. P_x : expectation, and SE_x : std. error of payoff, for offer x.

Figure 7: Responder behavior in the UG induces a binomial distribution of payoffs for each strategy chosen by proposers in the UG.

Variables	Model 1	Model 2	Model 3	Model 4
YNG	-5.391**	-5.414**	-6.024**	-5.859**
	(3.091)	(3.106)	(3.135)	(3.155)
$female$		1.079		0.657
		(3.107)		(3.151)
$gmatch$		-1.654		-7.913
		(3.271)		(6.637)
$foreign$			0.136	0.521
			(5.790)	(5.874)
$omatch$			-6.877**	-9.717**
			(4.144)	(4.923)
$gomatch$				8.270
				(7.632)
Constant	30.802***	30.855***	36.200***	38.166***
	(2.179)	(2.864)	(4.957)	(5.811)
Observations	181	181	181	181

Table 3: Linear regressions of RT as a function of the game and several control variables. (***) $p < 0.01$, (**) $p < 0.05$, (*) $p < 0.1$.

7 Pseudo-UG

The variety of choices in the YNG suggests the presence of pro-social types who play certain strategies regardless of their profitability. There may be correlation between RT and behavioral type in both the YNG and the UG. There are good reasons for wishing to eliminate such correlation from our data. For example, the hypothesis that larger RT corresponds to a higher level in a cognitive hierarchies model, may be falsely rejected only because behavioral types with a tendency to make proposal [50, 50], also tend to have a larger RT. Let us assume that the percentages of pro-social behavior observed in the YNG are representative of the percentages of pro-social "types" that are also present in the UG. To explain this better, suppose 35% of all proposals in the UG are [50, 50], while only 5% of proposals in the YNG are [50, 50]. According to our assumption, approx. 14% of the 35% of egalitarian proposals in the UG (5% of all proposals) are made by *egalitarian* types who would have made such a proposal under any circumstance. We wish to remove these proposals. Since we do not know which proposals among all [50, 50] proposals are made by egalitarian types, our procedure will randomly remove an appropriate percentage from the UG sample repeatedly, to simulate a data set that is behavioral-type free. Let f_i be the frequency of offer $i \in \{[70, 30], [60, 40], [50, 50]\}$ in the YNG. To generate the *pseudo-UG* data we randomly choose a fraction f_i of proposers in the UG who make offer i to be removed from the sample. We repeat this random process 100 times and obtain a sample with proposer offers and response times, which we call the *Pseudo-UG* sample.

The *Pseudo-UG* sample has mean offers that differ significantly (at the 5% level) between all pairs of speed categories (mean offer equals 38.09 for the slowest 25% proposers and 43.4 for the fastest 25%). Kullback-Leibler divergence is decreasing with RT. The histogram of offers is unimodal with mode at 50 for all speed categories except the slowest, which has two modes: one at 40 and one at 50. The pseudo-UG data maintains most qualitative features of the original UG data, with two differences: the relative frequency of proposal [90, 10] is higher overall in the pseudo-UG data and the slowest group has two modes instead of the unique mode at [60, 40] that we obtained with the original data. The hypothesis that fast behavior is more concentrated than slow behavior cannot be rejected with pseudo-UG data either. Also, the hypothesis that RT positively correlates with levels of reasoning is rejected with the pseudo-UG data as well as with the original data.