Interactive and Moral Reasoning: A Comparative Study of Response Times

Pablo Brañas-Garza
Debrah Meloso
Luis Miller

IGIER Working Papers 440

WORKING PAPER SERIES

Interactive and Moral Reasoning: A Comparative Study of Response Times

Pablo Branas-Garza, Debrah Meloso, Luis Miller

Working Paper n. 440

This Version: May 14, 2013

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Interactive and Statistical Reasoning: A Comparative Study of Response Times*

Pablo Brañas-Garza† Debrah Meloso‡ Luis Miller§

May 14, 2013

Abstract

When confronted with a new strategic situation individuals may use introspection to make a choice. We consider two frameworks to analyze introspection by positing that it can be either *interactive* or *statistical* in nature. Under interactive introspection agents construct an increasingly complex model of their opponent's rationality and choose an optimal response to this modeled opponent. This type of introspection is related to models of cognitive hierarchies or levels of reasoning. Under statistical reasoning, economic agents use introspection to recall behavior in other, similar, games and apply it to the current situation. Statistical reasoning relates to models of similarity-based learning across games. We run a laboratory experiment to explore whether subjects use interactive rather than statistical reasoning. We use response time (RT) and behavioral data from two different but related games (Ultimatum and Yes-or-No Game) to test the above hypothesis. We find no support for interactive reasoning but cannot reject the use of statistical reasoning by subjects in our large sample.

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

*A previous version circulated with the title ***Instinctive Response in the Ultimatum Game***. The authors gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation (grants ECO2012-30626, CSD2010-00034, and ECO2010-17049), the Government of Andalusia Project for Excellence in Research (grant P07.SEJ.02547), the Fundación Ramón Areces R+D 2011, 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

‡DEC, Università Bocconi, via Rontgen 1, 20136 Milan, Italy. **Corresponding author**. debrah.meloso@unibocconi.it

§BRIDGE & DFAE II, Universidad del País Vasco (UPV/EHU), Av. Lehendakari Aguirre 83, 48015 Bilbao, Spain

1 Introduction

We analyze behavioral and *response time* (RT) data from an experiment where subjects engage only once in one out of two possible strategic situations or games (ultimatum game and yes-or-no game). The general objective is to better understand how individuals face such unique strategic situations, which we will refer to as *one-shot games*. In particular, do individuals engage in introspection and if they do, what are valid models of the type of introspection they engage in?

The study of one-shot games is extremely 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 when it comes to their study in game theory, is that one cannot reasonably assume that agents will *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. Hence, if agents play Nash equilibrium in a one-shot game, then they must have an inner picture of how other agents play. We call this "introspection." This is different from interactions that are either natural or frequent (e.g., the crossing of a busy intersection by an experienced driver on her usual commute), where we expect experience or instinct to replace introspection.

The use of Nash equilibrium to study one-shot games is thus questionable unless one can reconcile it with a model by which agents can deduce or induce equilibrium behavior. Examples of such models are levels of reasoning and cognitive hierarchies, where agents iterate on a belief of rationality of other agents (Nagel [1995], Camerer et al. [2004]). Notice that in this class of models, agents have a picture of how others play that also justifies *why* they play in a certain way. An altogether different class of models on how one-shot games are played is based on *similarities*. Agents know how to play a new game because they have learned by playing similar games in the past (LiCalzi [1995], Mengel [2012], Jehiel [2005], among many others).[1]

Our analysis of introspection focuses on two frameworks based on the two sets of models mentioned above. We refer to them as the *interactive* (or model-based) framework and

[1]Work in this area does not address how an agent faced with a new game chooses the similarity group relevant for this game, which is our concern. These theoretical models are therefore, only related to our approach, but do not formalize it. As discussed in Grimm and Mengel [2012], agents can treat similar games in different ways: by adopting equal beliefs, equal best-response correspondences, or even equal actions across them. This distinction is relevant to our findings and we briefly discuss it in section 4.

the *statistical* (or empirical) framework. The former framework is formalized either with a cognitive hierarchies model (with appropriate assumptions on level-0 beliefs; a levels of reasoning model additionally requires assumptions on the resolution of indifference) or with levels of backward induction. Intuitively, it posits that subjects best respond to an opponent whose rationality is defined by her own best responding to an analogously-defined rational opponent. More thought, in this case, implies more levels in the mutual assessment of each other's rationality by opponents. The latter, statistical, framework does not require subjects to postulate a model of rationality of their opponent or themselves. Instead, thought is dedicated to associate the new, unique game, to past experiences. In this way subjects can construct a growingly accurate prediction of the profitability of each available choice, probably via an accurate prediction of play by the opponent. This framework relates to the class of *similarity-based* models in that it posits that experience with similar games guides behavior in one-shot games. Notice that both frameworks posit that agents engage in strategic reasoning. The question is how this reasoning takes place.

The growing literature in experimental economics using RT measurement has given much attention to games that have a social aspect (Rubinstein [2007], Piovesan and Wengstrom [2009], Rand et al. [2012], Fischbacher et al. [2013(in press)], among others).[2] We also choose such games. We use two games and our analysis is performed across three dimensions: that of the game, that of response time, and that of behavior. The two considered games are the *Ultimatum Game* (UG) and the *Yes-or-No Game* (YNG), both well-studied in the experimental economics literature (see Guth [1995] for a literature review until 2000, other studies have since followed). Players in either game can take on the role of proposer or of responder. In their setup and description, the two games differ only in what responders are allowed to do. This permits a very clean comparison of response times of proposers, since differences cannot be attributed to sensorial or motoric activities performed by proposers when choosing in either game: the interfaces and choice sets are identical (see figures A.2 and A.3). On the other hand, the two games have very different *strategic* properties while maintaining similar social and emotional connotations.

The way in which we use RT requires the assumption that RT is at least positively correlated with a choice by agents to engage in deeper introspection before making a decision in the game. Agents with different RTs are assumed to engage in more or less steps of the same introspective process. This sets our work apart from much of the previous literature on RT in experimental economics, where longer RTs are associated with the

[2] It is not free from consequences to choose such games. Although by positing that RT and game combine to produce behavior (as opposed to RT on its own producing behavior of a certain type) we side-step some of the consequences of game choice, other implications remain. We return to this point in section 4.

suppression of instinctive reactions that do not stem from introspection (e.g., fast responses are pro-social while slow responses are reasoned, as in Rand et al. [2012]). This assumption carries over to our across-games analysis: agents in either game use the same form of introspection which may lead to different behaviors because the situations (inputs to the process) differ.

When combined with the games played by our subjects, the two frameworks we consider lead to specific questions. Firstly, we ask whether RT is on average equal across games. If it were, it would be questionable whether any introspection takes place at all! Secondly, we ask, along the lines of the interactive framework, whether proposers in the UG that have larger RT also make offers that are closer to subgame-perfect equilibrium behavior (offer the smallest possible share to the responder). Finally, for the statistical framework we use responder behavior as the reference for empirically-accurate behavior and ask whether proposers in the UG that have larger RTs come closer to this empirically-accurate behavior. We find no support for the interactive framework while the statistical framework stands strong within our data. Further results are presented to check the robustness of our findings and dig deeper into what exactly "statistical" subjects recall and assess with introspection: the distribution of opponents' behavior, a prescription for own behavior, or their own willingness to try out (strategically) risky offers.

When compared to the small but growing literature measuring RTs in experimental economics, our work is the first to use the comparison across two different situations to elucidate the mental processes that underlie decisions in a strategic setting. Like in Rubinstein [2007] and Arad and Rubinstein [2012], we focus on a one-shot interaction in a novel situation. Coricelli and Nagel [2009] test a null hypothesis of conscious introspection in the presence of human partners in a game, which closely relates to our first hypothesis. In their study, RT is only a marginal measurement, while brain activity measurement via fMRI is central. Arad and Rubinstein [2012] use subjects' data from participation in two different games in order to cross-check subject *types* and – also related to our work – they hypothesize a specific generalized levels model of reasoning for participants in their games. They do not, however, compare both behavior and RT across games. The latter is a widespread strategy in traditional RT studies in experimental psychology (see, for example, ch. 3 of Jensen [2006]) and is also used in Bergert and Nosofsky [2007] to perform a horse race between two proposed models of individual decision making in individual (as opposed to social or strategic) situations. Their work is relevant in that it provides a bridge between the classical studies in Psychology that consider simple decisions and differences in RT measured in milliseconds, and the highly complex, interactive decisions considered in Game Theory. Also Rubinstein [2012] focuses on individual decision problems. He finds large behavioral differences along RT categories, with a large number of "mistakes" among fast subjects.

Our finding that longer RTs in both games come along with a higher dispersion of behavior relates to Piovesan and Wengstrom [2009], Greene et al. [2001], and Knoch and Fehr [2007], among others. Piovesan and Wengstrom [2009] have subjects repeatedly face a situation with identical strategic properties but different levels of *equality* and no fully-egalitarian choice. Their setup is one designed to create moral dilemmas with minimal strategic concerns (dictator games), and they find that indeed a higher difficulty of the dilemma leads to larger RTs. Unlike in our work, in Piovesan and Wengstrom [2009] introspection about norms in their setting has no strategic consequences. Greene et al. [2001], on the other hand, focus on the role of emotions in the creation of moral dilemmas. Their main finding is that moral problems with high emotional charge in a given direction lead to instinctive actions in that direction, but very long RTs to overcome instinct and choose in the opposite direction. At the same time, no such correlation between RT and action was found in the absence of emotional charge. The manipulation of emotional charge is absent from our work as well as that of Piovesan and Wengstrom [2009], but importantly suggests another factor that may affect individual behavior in novel strategic settings. Fischbacher et al. [2013(in press)] uses mini-ultimatum games to classify subjects according to the prevalence of social motives in their decision making. It relates to our work only in their finding that larger RT is associated with empathy, which in turn is suggestive of the use of a levels-of-reasoning framework by subjects (however, our data do not support such a model). Also related to pro-social attitudes and RT is work by Rand et al. [2012] using public-goods games. Finally, Knoch and Fehr [2007] relate reasoned self-control with (responder) behavior in the UG, arguing that more self control is required to behave less selfishly. This study does not involve RT measurement but the impairment of conscious brain processes associated with self-control. A translation of such conscious processes into longer RT does not survive the scrutiny by data in our experiment, since it is not selfless behavior that prevails with longer RT, but rather rarer choices and a larger variety.

In the next section we present the experimental design and hypotheses. Section 3 presents the results and section 4 discusses them and concludes.

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]). Full experimental instructions can be found in Appendix A.1. 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. No communication between subjects was permitted. Subjects could not identify which other participant they were interacting with. 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 other half (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 (see figure A.2 for a screen shot of responders' choices in the UG). 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 proposal 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

[3]We used neutral names, X (proposer) and Y (responder), for the two roles. Although partners were anonymous, subjects were informed of the gender of their partner. Subjects' RTs are indistinguishable across gender (t-test and Kolmogorov-Smirnov test with p-values of 0.959 and 0.428, respectively).

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

this share the *Minimal Acceptable Offer* (MAO). In our experiment we have a MAO for all but four responders and we use it – instead of the entire strategy description – for our analyses.[5]

In the YNG, the responder, without knowing the proposal of his partner, was asked to choose whether he would accept every offer or reject every offer from the proposer (see figure A.3 for a screen shot of responder choices in the YNG). Hence the name "Yes" or "No" game, since the responder can either say yes to every proposal or no to every proposal. 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 offer. We will use bracketed pairs to denote the available offers: [90, 10], [80, 20], [70, 30], [60, 40], [50, 50], 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.

It is important to point out the differences between the two considered games:

Responder strategy sets. Responders in the UG had a large strategy set (larger than in the YNG). A single strategy is a list specifying an *accept* or a *reject* for each of the nine possible proposals. Without the assumption of monotonicity, this means that they had 2^9 possible strategies. Responders' strategies were monotonic if they had a MAO, which reduced the set of strategies to only 9 (MAO equal to each one of the possible proposals). Even in this latter case the strategy set of responders in the UG was larger than that of responders in the YNG. In this latter game, responders had only two possible strategies: "Accept any offer", or "Reject every offer". Notice that this difference in strategy sets across games implies that also the computer interface and information input faced by responders in our experiment differs across games.

Proposer strategy sets. Proposers in the two considered games had **identical** strategy sets. In either game, the proposer had to decide which one out of nine possible proposals to make to the responder. In terms of implementation, proposers in our experiment faced the exact same interface and information input regardless of the game they participated in. This is an important reason behind our choice of games to compare (UG and YNG)

[5]We do not use the data of the responders who have non-monotonic strategies (no MAO). The four discarded cases 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.

and our choice to focus analysis on behavior and response times of *proposers*.[6]

Equilibrium Predictions. There are two layers of difference between the equilibrium predictions of the two games. On one hand, the equilibrium in the YNG is unique, while there are multiple Nash equilibria in the UG. On the other hand, the equilibrium in the YNG is an equilibrium in weakly dominant strategies, while none of the equilibria in the UG has this property. The unique, dominant-strategy equilibrium in the YNG has the proposer offering [90,10] and the responder accepting every offer (recall that we have set up the games so that 90 is the largest amount that the proposer can keep to himself – offers of [100,0] are not allowed) . In the UG, every strategy profile where the proposer offers $[100-x,\ x]$, and the responder sets a MAO equal to x, is a Nash equilibrium of the game. Attention is often focused on the unique *subgame-perfect* equilibrium (SPE) of the UG, which has the proposer offering [90,10] and the responder setting his MAO at 10.[7] The implication of these equilibrium predictions is that even though proposers have identical strategy sets in these two games, their choice of strategy may require completely different considerations.

A proposer in the YNG may ask herself what her payoff of choosing a given strategy is when the responder chooses *accept any offer* or when he chooses *reject every offer*. However, once this proposer realizes that she has a dominant strategy, she no longer faces the need to form a belief about what the responder is likely to do! Notice that unlike in a dictator game (see, e.g., Piovesan and Wengstrom [2009]), the proposer's payoff depends on responder behavior but, crucially, she can quickly realize that she need not reflect about the likelihood of each responder behavior in order to figure out her own best interest (dominant strategy). In the absence of a dominant strategy, the proposer in the UG cannot escape the need to reflect about responder behavior. Moreover, in spite of the unique SPE, we argue that the proposer in the UG faces a real (strategic) risk that the responder state a MAO different from 10, without needing to assume an irrational responder (other MAOs can be sustained in equilibrium). In particular, a proposer who makes an offer $[100-x,\ x]$ risks obtaining nothing if the responder sets a MAO larger than x, or obtaining "too little"

[6]It is important to stress that the simple fact that responders have a larger strategy set in the UG does not imply a *more difficult* strategic situation for proposers and, moreover, it does not change the interface for choice faced by proposers. It is the nature of the difference in strategies of responders across games and not the mere size, that may make a difference for proposers. A trivial example would be to consider a variant of the YNG where responders had 7 additional payoff-irrelevant strategies: the size of the strategy set would become the same as in the UG with monotonic strategies, but the strategic nature of the game would remain unaltered.

[7]The game we present to our subjects is the strategic form of the sequential game where the proposer first makes an offer and the responder next chooses whether to accept or reject this offer. The mentioned SPE corresponds to this sequential game.

if the responder sets a lower MAO. Hence, in the UG the proposer needs to formulate a belief about the responder's behavior, and formulating such a belief is difficult.

Both games have subjects that mutually affect each other's payoff and are therefore subject to considerations of guilt, equality, or other social feelings (for a survey of such considerations in game theory, see Attanasi and Nagel [2008]). Social feelings that are intrinsic are maintained in a dictator game as well, but those feelings involving expectations and surprise (either positive or negative) toward the other's actions can only be considered in the context of games where both players take actions. Moreover, in a dictator game the proposer (dictator) takes on full responsibility over the well-being of the responder (passive subject), potentially giving rise to an entirely different class of social feelings and norms. Finally, the wording used to describe the game to proposers (main subjects of our analysis) is quite homogenous across the YNG and UG: in both cases the context is described as one of negotiation with choices for both subjects. This feature would be lost if a dictator game instead of a YNG were used. To summarize, we choose our games so to make them clearly different in terms of the degree of interactive thought they demand from optimally-behaving subjects, but minimizing differences both in experimental implementation and social character of the games.

2.2 Response Times

Subjects participate in either game via an anonymous electronic interface. We record subject strategy choices and payoffs, but also the time elapsed between the moment in which they are presented with the problem and the moment when a choice is made, which we will call the *response time* (RT).[8] Obviously, this time frame depends on the way in which the game is presented and the physical motions subjects must make in order to state their choices. The absolute numbers, hence, bear little meaning. Instead, it is the comparison between the two games and between groups of subjects who play differently in the same game that will matter. We make the problem presentation and choice-making motions as similar as possible between games, to ensure a meaningful comparison of RTs (Appendix B, figures A.1, A.2, and A.3).

The measurement of RT for decisions in the UG replicates the work of Rubinstein [2007], but in a standard laboratory setting. Subjects in our experiments are compen-

[8] 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, in that sense, very different. Measurement is less precise and the economic models we use as reference are not explicitly designed to uncover mental structure.

sated and controlled for communication and distractions. Rubinstein [2007] reports data collected from several thousand subjects who partake in a number of game-type situations. Participants' responses were collected from an on-line survey experiment and were not incentivized. The arguments for and against these two methodological choices are clear. On the one hand, an on-line non-incentivized experiment is cheaper and provides a considerably larger sample. Standard economic experiments are more expensive and use much smaller samples. However, total control over the subjects' decision process is crucial for studying RT, and control is definitely the main methodological feature of laboratory experiments.

Having said this, it is important to emphasize that the response time measurement we make is not as precise as in standard decision experiments in psychology. Moreover, subjects are not incentivized for speed and are not aware that their RT is being measured in the experiment. However, presumably subjects do have an opportunity cost of their time and will naturally trade off speed of decision and "correctness" of their choice, as they would do in standard RT experiments where this trade off is induced. Still, this consideration stresses even more the comparative, as opposed to absolute, meaningfulness of the RT data we collect and analyze. We will be interested in the interaction between subjects' RT, their choices, and the game they are faced with.

2.3 Experimental Questions

In our first hypothesis we stipulate that RTs of proposers in the UG are larger than those of proposers in the YNG (in distributional terms, proposer RTs in the UG first-order stochastically dominate those in the YNG). We focus on proposers because they have identical strategy sets and decision interfaces across games, which allows us to focus on the strategic reasons for differences in RT.

With this first hypothesis we wish to establish whether subjects naturally recognize the strategic intricacies of either game and react by spending time in the game where it matters. A methodology where subjects are given an a priori time limit would not have served this purpose. In this first hypothesis we focus on subjects' choice to dedicate more time to a situation that is more complex in a strategic sense.

Our second set of questions asks what it is that our subjects invest time in. We focus on the interaction between behavioral and RT data in the UG, using the YNG mostly as a benchmark. We use two alternative frameworks to guide our analysis: The *interactive* and the *statistical* framework.

Interactive Reasoning

The first framework is one where subjects deduce the strategy their partner is likely to play from an ever more sophisticated model of the partner's rationality. In other words, proposers construct a model that does not only predict, but also *justifies* the strategy they expect their partner to play. This framework leads to the formulation of our second hypothesis stating that slower proposers in the UG make lower offers since they tend toward play of the SPE of the game.

This hypothesis can be derived using either a model of cognitive hierarchies (Camerer et al. [2004]) or one of levels of reasoning (Nagel [1995]). The former requires milder assumptions in order to derive the hypothesis, hence we intuitively summarize it here. We assume that subjects are in principle able to engage in many levels of reasoning but require time and effort in order to do this. A zero-level proposer is assumed to make any offer between 10 and 50 (responder's share) with equal probability. A zero-level responder is assumed to set a MAO between 10 and 50 with equal probability. Level one players best-respond to zero-level opponents. Level two players best-respond to a combination of level zero and level one players, etc.[9] Proposers of level one find it optimal to offer 50, since level zero responders are likely to set a high MAO. Responders of level one or higher find it optimal to set a MAO of 10. This means that proposers face a risk of rejection only by level-zero responders. Their effect on a proposer's expected payoff becomes smaller with the level of reasoning, making lower offers, and finally, the lowest offer, more profitable.

A levels-of-reasoning model, where a level k player best responds to level $k-1$ players only, also achieves the same prediction, but additional assumptions must be made about the way in which responders resolve indifference between several equally-profitable strategies. Finally, one can also justify this hypothesis with a simple model based on backward induction. Under this model a fast proposer makes no inference about responder behavior, while a slower proposer dedicates thought to deduce the responder's best action at some or all of his decision nodes. In particular, a slow proposer deduces that the responder's optimal action after offer [90, 10] is to "accept." Clearly, such a model converges to SPE behavior.

This framework thus translates to the simple question of whether large RT means that offers become less dispersed, with a mode at [90, 10].

[9] Camerer et al. [2004] assume a Poisson distribution of players of each level. A level k player assigns zero probability to players of levels k and higher, and correctly normalizes the frequency of lower-level players conditional on this (wrong) belief. For our purposes, any distribution giving positive probability to levels 0 and 1 yields the same behavioral prediction.

Statistical Reasoning

A second framework envisions proposers who use time to evaluate the empirical profitability of different choices. They can do so by recalling how they and other potential subjects behaved in similar situations. Thought can be dedicated to identifying the relevant similarity context for the game at hand and to recalling what behavior was either performed or considered appropriate in such similar games. There are different possibilities about what exactly agents retrieve: the norm of behavior prevalent in the game's similarity context and, hence, a prescription on how to play; or the empirical distribution of play by opponents in similar situations, and hence, a distribution against which to best respond. In either case, we postulate that empirical accuracy improves with RT.

We thus ask whether the offers of proposers with larger RTs, on average, are more profitable given the distribution of MAOs we observe among responders in our experiment. We do take some steps in the direction of asking whether they are more profitable because slow proposers better estimate this distribution of MAOs or because they discover the appropriate norm by which to behave. However, results on this regard are only suggestive, motivating future work.

3 Results

As hypothesized, overall RTs are larger in the UG than in the YNG, for both proposers and responders. Here we will mostly focus on proposer data. The mean proposer RT in the UG is 30.94 seconds, while it is 25.28 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 0.0323). Moreover, a non-parametric test of stochastic dominance of the distribution of RT for the UG over that for the YNG (Kolmogorov-Smirnov test) has a p-value of 0.015. If we separate subjects by offers, even though the comparison is not always significant, due to small samples, it always is the case that RTs are on average smaller in the YNG than the UG. Figure 1 (all three subplots) illustrates these results.

3.1 Interactive Reasoning

We now turn to the second of our questions, focusing first on what we have called *interactive reasoning*. For the YNG we expect to see proposers in all speed categories making the same, equilibrium offer of [90, 10]. For the UG we expect to see a change from [50, 50] toward [90, 10] (see section 2.3). A different visualization of our results will help in dealing with this question. In figures 3a and 3b we look at the offers made by proposers located in different *RT* categories. Following Rubinstein [2008], the categories are determined with respect to the in-sample distribution of *RT*s. The 10% fastest proposers in each game are grouped in category *Very Fast*, and the 10% slowest are grouped as *Very Slow*. The remaining 80% of proposers is divided in two equally-sized groups: the *Slow* and the *Fast* group.[10]

Figure 2a clearly indicates that the most frequent offer in the UG is [50, 50], and figure 3a indicates that this is overwhelmingly the modal offer among the *Very Fast* proposers. As *RT* increases, the distribution of offers changes significantly, but not toward a unimodal distribution with mode at [90, 10]. Instead, the distribution of offers of *Very Slow* proposers is much more dispersed than that of *Very Fast* proposers. We use the Kullback-Leibler divergence from a uniform distribution to measure the concentration of the distribution of choices for different *RT* categories. The smaller the Kullback-Leibler divergence, the more dispersed the distribution. For the UG this divergence decreases steadily as we move from the *Very Fast* to the *Very Slow* category, meaning that each "slower" category is more dispersed than the previous one.[11]

Although the mean offer of the *Very Fast* proposers is significantly larger than that of other speed categories, the mean offer remains high even for the *Very Slow*, and pairwise comparisons between all other categories (excluding the *Very Fast*) are not significant (from lowest *RT* to highest *RT*, mean offers to responders are 47.78, 40.54, 40, and 38.39). Furthermore, there are no offers of [90, 10] among the *Very Slow* proposers.

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 2b and 3b).[12] When compared to behavior in the UG, $t-$tests of equal mean offers are rejected at the 1% level in favor of the alternative hypothesis that offers are larger in the UG, both for the entire sample and for *Very Fast*, *Fast*, and

[10]The categories differ significantly in the range of *RT*s they include. The *Very Fast* category is very narrow, suggesting an immediate, instinctive reaction. The dispersion, as measured by the range and the coefficient of variation of *RT* within each category, increases monotonically as we move through slower categories. This is the case for proposers and responders in both games.

[11]Let f_i be the frequency of offer i and let $q_i = 1/5$ be the probability of each offer under the uniform distribution. The uniform is our benchmark for *maximum dispersion*. Then, the Kullback-Leibler measure is

$$KL(f,q) = \sum_{i=1}^{5} s_i, \text{ where } s_i = \begin{cases} f_i \ln\left(\frac{f_i}{q_i}\right) & \text{if } f_i > 0 \\ 0 & \text{if } f_i = 0 \end{cases}.$$

The measure is smaller when the distribution of offers in our data is more similar to a uniform distribution, meaning that it is more dispersed or less concentrated.

[12]This result contrasts with the finding that pro-social behavior is instinctive in games (see Rand et al. [2012]). Not surprisingly, the growing literature on *RT* in games is showing that what is "instinctive" depends on the game that is used in the experiment.

Slow proposers. Pearson's χ^2 test comparing the distribution of offers across games rejects the null of equality at the 1% level for the entire sample and for speed categories *Fast* and *Slow*, and at the 5% level for category *Very Fast*. For *Very Slow* proposers the mean offers get closer (equality is still rejected at the 5% level of significance) and we can no longer reject the null hypothesis that offers in both games come from the same distribution (Pearson's χ^2 with $p-$value of 0.15). This latter result agrees with the biggest similarity we find between the two games: in both the UG and the YNG the distribution of offers becomes more dispersed for larger *RT*s. Table 1 shows the Kullback-Leibler divergence indices that corroborate this finding.

All above results qualitatively follow through if we consider an alternative categorization of *RT*s into approximately equally-sized groups, thus avoiding the problem that hypotheses are rejected in some cases and not in others simply because sample sizes differ. We consider a division into 5 speed categories containing approx. 20% of proposers each – we refer to it as the *quintile* categorization. Also for this categorization dispersion increases with *RT*, and the above-mentioned comparison with the YNG (by speed categories) still holds through.

The main lessons to be drawn from the results mentioned so far are that slower proposers in the UG are not more likely to propose [90, 10] than the faster proposers and that, instead, as *RT* increases, so does the dispersion of offers made by proposers in the UG. The correlation between dispersion of choices and *RT* is also true of offers made in the YNG and of MAOs of responders in the UG (we will return to this result in the following section).

Pseudo-UG

The variety of choices in the YNG suggests the presence of "irrational" or "stubborn" types who play certain strategies regardless of their profitability. We will later ask whether this is a valid conclusion. Here we take it as true and check whether results in this section are robust to the presence of such types. We modify the distribution of offers in the UG using YNG choices as reference: since we are interested in the way in which *RT* invested in thought affects behavior, we wish to remove from our sample subjects who play certain strategies regardless of their profitability. 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. The mean offer of the *Very Slow* is still significantly different from the lowest, [90, 10] offer (mean offer equals 39.71 for the *Very Slow* proposers and 37.92 for the fifth quintile). Kullback-Leibler divergence is decreasing with RT: like for the real UG data, the *Slow* category or category IV in the quintile categorization is more concentrated than its neighbors, but this concentration is driven by offers of [50, 50] and [60, 40]. Hence, even after adjusting for the presence of "stubborn" types, offers do not converge to being unimodal with a mode at [90, 10] (figure 4a).

3.2 Statistical Reasoning

The second framework we use to analyze the interaction between behavior and RT postulates that subjects use RT in order to make a better assessment of the empirical distribution of their partners' strategies, based on their experience in *similar* environments, or more simply, on norms of behavior that are acceptable in their environment.[13] Since experience and norms are unknown to us as experimentalists, we measure the extent to which this framework has validity in our experiment using the realized distribution of *responder* plays as a proxy for these norms. We thus focus on the distribution of payoffs that is generated by the distribution of responder choices, for each offer that a proposer can make. Figure 5a shows the histogram of responder MAOs.

Let g_i denote the fraction of all responders with MAO equal to $i \in \{10, 20, 30, 40, 50\}$. A proposer making offer $[100 - x, x]$ has a probability $G_x = \sum_{i \leq x} g_i$ of having his offer accepted, and $1 - G_x$ of having it rejected. Thus, a proposer offering $[100 - x, x]$ gets payoff $100 - x$ with probability G_x and 0 with probability $1 - G_x$. Table 5b shows the probability, G_x, expected payoff, P_x, and standard error, SE_x, of the payoff associated with each possible offer, x. Offer [60, 40] has the highest expected payoff and also has lower variance than all smaller offers.

We use the information on expected payoffs to each strategy to look at the distribution of expected payoff of proposers in different speed categories. Clearly, this distribution will be a simple re-organization of the histogram of offers of each speed category: proposers offering [80, 20] have the lowest expected payoff, followed by proposers offering [70, 30], and so on, until the offer [60, 40], with the highest expected payoff. For easier visualization, figure 6a shows empirical distribution functions (smoothed histograms) for all

[13] The use of *norm* here relates to Bicchieri [2006]'s definition of a social norm. Bicchieri [2006] defines a social norm as a behavioral rule that fulfills two conditions: individuals involved in a situation know of the existence of the rule and they prefer to conform to it conditional on their empirical (first-order) and normative (second-order) expectations.

speed categories in a single graph. Table 2 has the basic statistics of these expected payoff distributions divided in *RT* categories.

The *Slow* category has the distribution with largest mean, median and mode (it is shifted to the right of the other speed categories). The frequencies of each payoff among *Slow* proposers is significantly different from that of *Very Fast* and *Very Slow* proposers at the 10% level, but not from that of *Fast* proposers (Pearson's χ^2 test). When we consider the equally-sized speed categories, the fourth quintile has a distribution of expected payoffs shifted to the right with respect to all other categories, but differences are never significant.

Finally, figure 6b shows the empirical pdf of realized payoffs by UG proposers in different *RT* categories. Again, differences among speed categories are not significant. When it comes to realized payoffs, the *Very Slow* group has the highest mean, but the highest median and mode still pertain to the *Slow* (in the quintile categorization, the fourth quintile has the highest mean, median, and mode). As figure 6b shows, payoff variance increases with *RT*, which is consistent with the higher dispersion of offers.

Summing up, we have so far seen that, given the distribution of responder MAOs, the offer [60, 40] has the best mean-variance properties. *Slow* (or fourth quintile) proposers modally make this offer, which gives them a distribution of expected payoff that dominates that of other *RT* categories, as well as a higher median and mode of realized payoffs. Following our line of thought, our experimental results indicate that slower proposers do a better job at guessing the distribution of MAOs. Does this guessing happen in a conscious way or is it mediated by norms? The following results are concerned with this question.

As already mentioned, also in the YNG the dispersion of offers increases with *RT*. Clearly, in the YNG this leads to a worsening of the distribution of expected and realized payoffs, since [90, 10] is a weakly dominant strategy. We hypothesized that large offers in the YNG could be due to the presence of *stubborn* types when we generated the *Pseudo-UG* sample. The problem with this hypothesis is the choice of the stubborn types. About 52% of all proposers in the YNG choose offers different from [90, 10], and among these, the largest fraction (approximately 35%) choose offer [60, 40], followed by [50, 50] (approx. 30%), and in third place [70, 30]. These offers are, respectively, the offer with highest expected payoff (and second-to-lowest variance), the offer with lowest variance (and third-best expected payoff), and the offer with second highest expected payoff in the UG. It follows that the *Pseudo-UG* sample shows a lower expected profitability of *RT* than the real UG sample. *RT* categories in the *Pseudo-UG* sample have distributions of expected payoff that are more similar across categories because the faster subjects do less well on average. In other words, our "stubborn" types do well in the UG!

Furthermore, the relation between *RT* and *stubborn* behavior in the YNG is similar to that in the UG: [50, 50] is replaced with [60, 40] and the even riskier [70, 30] as *RT* increases. The picture that is emerging is one where our subjects, with increased *RT*, adopt (strategically-risky) norms that are highly profitable in the UG, and a fraction of subjects do so even when faced with a "similar" setup where these norms are no longer profitable.[14]

Results on UG responder behavior consistently add to this picture. Figure 7 summarizes the results obtained on responder behavior in our experiment.[15] The distribution of MAO for the *Slow* and *Very Slow* responders is less concentrated than for the faster subjects and there is a move away from the weakly dominant strategy, toward more intermediate strategies (Kullback-Leibler divergence is much lower – approx. half – for category *Very Slow* than category *Very Fast*). In other words, it is the slower subjects that generate the distribution of MAOs such that the offer [60, 40] becomes more attractive than offer [90, 10] for proposers. It is not individually optimal for any responder to set a MAO above 10, but it collectively induces a proposer behavior that improves all responders' well-being. Notice further, that like for proposers, *Very Fast* behavior favors the strategy with 0 strategic risk. As *RT* becomes larger, subjects favor strategies that are risky and that pay off only *given* the realized behavior of subjects in the complementary role. No amount of time makes subjects favor behavior that pays off given an assumption of individual rationality at every step by subjects in the complementary role.

These last results suggest that subjects use time to understand the similarity group to which a new game belongs and retrieve norms to play such games. This, as opposed to retrieving a belief of behavior by opponents and best responding to it. Returning to the models motivating the statistical-reasoning framework, and the taxonomy used in Grimm and Mengel [2012], our results suggest that players fix *behavior* across games they classify as similar (as opposed to either beliefs or best-response correspondences only).

[14]Notice, however, that subjects do not go from the zero-risk [50, 50] offer to the risky, but subgame-perfect offer [90, 10]. Hence, one cannot simply argue that *RT* reduces risk aversion of subjects, pushing them toward SPE behavior. Instead, it does reduce their risk aversion and push them toward the empirically-optimal strategy in the UG.

[15]As expected, due to a mixture of reasons, responders in the YNG have uniformly lower *RT*s than in the UG. Average *RT*s are 40.06 in the UG and 15.22 in the YNG, statistically different from each other (t-test with p-value< 0.001). Non-parametric tests corroborate the finding that *RT*s are lower for the YNG than the UG.

4 Discussion and Conclusion

We start from the premise that there is no such thing as *intuitive behavior*, since behavior is irremediably situation dependent, even if the process used to generate behavior is unique. This is not a philosophical statement but an operational one: an approximate algorithm may generate an exact solution or a very large error, depending on the parameters of the problem instance that it is asked to solve. This does not mean that response times, as a tool for research in social science must be discarded. To the contrary. Since the identification of processes is harder than that of behavior, an added dimension of measurement is very welcome.

We focus on one-shot games because there are strong reasons to expect convergence to equilibria (steady states) when an experience is repeated, but we do not know what to expect when an experience is faced only once. We expect our subjects to use past experience in similar situations, induction and other forms of introspection to choose their behavior. To guide our data analysis, we focus on two alternative frameworks about what type of thinking subjects engage in. We call them interactive and statistical reasoning.

Firstly, we find strong evidence that subjects are *selective* about investing time in one or other situation they are faced with. More time is spent, on average, in the situation that is strategically more complex. This justifies our follow up questions, since it is sensible to believe that subjects with larger RTs are aware that the profitability of each of their actions depends on the behavior of others and, hence, they must reflect about how they *should* play.

We find no support for the interactive reasoning framework. This framework postulates that thought is dedicated to forming an ever more sophisticated image of one's partner's rationality. It is one extreme of the spectrum: if launched on a strange planet, a proposer in a UG would need to build from zero a prediction of alien behavior. If this proposer knew the alien's utility function, he could combine it with an assumption that aliens want to maximize utility to build this prediction. Under this framework we expect slow proposers, who have a fairly sophisticated model of their partner's rationality, to make the lowest offer, [90, 10].

Nonetheless, the distributions of offers of proposers and that of MAOs of responders match up and they do so thanks to the behavior of slower subjects! Proposers start out from playing a zero-risk strategy, [50, 50], for which there is no need to outguess responders, but with time they start taking risks that are justified given the distribution of responder MAOs: slower proposers modally play the offer with the highest expected payoff (and lowest positive variance). Responders, on the other hand, start out playing the zero-risk strategy, MAO= 10, which is also weakly dominant, but with time they start taking risks. Their risks are never justifiable on the basis of individual profit, but it is these responders that sustain the – beneficial for responders – observed behavior of proposers.

An important weakness of our analysis is that it is *across-subjects*, not within-subject, but it implicitly ignores that subjects may differ in the speed at which they reason. As a matter of fact, when we analyze the distributions of expected payoff of the real and the *Pseudo-UG* samples, we find that *Very Slow* subjects do worse than *Slow* subjects. This is consistent with a model where there are two sources of variance. On one hand, subjects differ in their perceived benefit of thinking and therefore choose different amounts. On the other, subjects differ in the productivity of time in generating "thought." Subjects who know they don't understand the game may choose to think before making a choice but they nonetheless make a low profit choice because they do not make a good use of time (or start at a very low level of understanding). We focus on the former source of heterogeneity and disregard the latter. Notice, however, that if all heterogeneity in our data came from differences in time productivity (ability to think), there would be no systematic behavioral differences between *RT* categories. Hence, we feel comfortable that our results are not driven by our ignoring the existence of heterogeneity in ability across subjects. Work by Fischbacher et al. [2013(in press)] and Arad and Rubinstein [2012] partially address the relation between *RT* and exogenous types.

A final warning is in place regarding results being game specific. Our finding that statistical reasoning plays an important role when subjects face one-shot games helps understand previous results where, in specific types of games subjects were altruistic when fast and less so when slow (Rubinstein [2007], Rand et al. [2012]) or vice-versa (Piovesan and Wengstrom [2009]). Behavior of course depends on the game at hand and, even more, on the games individuals consider *similar* to it. These similar games may, as in the case of the YNG, in fact have a very similar framing but a very different strategic nature than the game at hand. Public-goods games (as used in Rand et al. [2012]) are a clear example, since real-world public-goods games – unlike laboratory voluntary-contribution games – usually come with a punishment for not contributing. By moving the focus away from the specific behavior and toward the type of introspection, our approach escapes this problem of game specificity. But not completely! Our subjects may in fact choose to engage in statistical thinking because they quickly perceive the social nature of the game they are faced with and its potential similarity with other, social situations they have experienced. If, instead, the game is deprived of social content or is truly unfamiliar, subjects may find it more desirable to engage in interactive reasoning. We do not address this question here and leave it for future research.

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.

Giuseppe Attanasi and Rosemarie Nagel. A survey of psychological games: Theoretical findings and experimental evidence. In Alessandro Innocenti and Patrizia Sbriglia, editors, *Games, Rationality and Behavior. Essays on Behavioral Game Theory and Experiments*, chapter 9, pages 204–232. Palgrave, McMillan, Houndmills, 2008.

Bryan Bergert and Robert Nosofsky. A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology*, 33(1):107–129, 2007.

Cristina Bicchieri. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, 2006.

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

Giorgio Coricelli and Rosemarie Nagel. Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Science*, 106(23): 9163–9168, 2009.

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*, 2013(in press).

Joshua Greene, Brian Sommerville, Leigh Nystrom, John Darley, and Jonathan Cohen. An fMRI investigation of emotional engagement in moral judgment. *Science*, 293:2105–08, 2001.

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.

Veronika Grimm and Friederike Mengel. An experiment on learning in a multiple games environment. *Journal of Economic Theory*, 147(6):2220–2259, 2012.

Werner Guth. On ultimatum bargaining experiments – a personal review. *Journal of Economic Behavior and Organization*, 27(3):329 – 344, 1995.

Philippe Jehiel. Analogy-based expectation equilibrium. *Journal of Economic Theory*, 123 (81–104), 2005.

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.

Marco LiCalzi. Fictitious play by cases. *Games and Economic Behavior*, 11:64–89, 1995.

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

Friederike Mengel. Learning across games. *Games and Economic Behavior*, 74(2):601–619, 2012.

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.

Ariel Rubinstein. Response time and decision making: A “free” experimental study. Available from author, 2012.

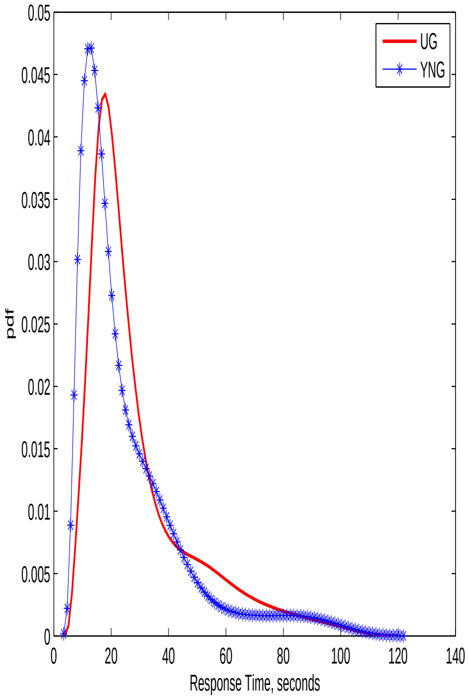
Reinhard Selten. Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperiments. In Heinz Sauermann, editor, *In Beiträge zur experimentellen Wirtschaftsforschung*, volume I, pages 136–168. J.C. Mohr (Siebeck), 1967.

	Very Fast	*Fast*	*Slow*	*Very Slow*
UG	1.078	0.396	0.424	0.394
YNG	0.460	0.529	0.173	0.240
Pseudo-UG	*1.164*	*0.443*	*0.386*	*0.382*

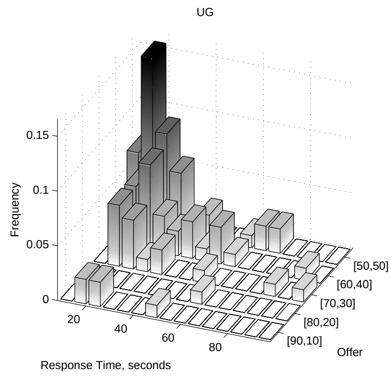
Table 1: Kullback-Leibler divergence from a uniform distribution over all offers. Lower values correspond to a higher dispersion of offers.

	Median	Mode	Mean
Very Fast	50	50	51.94
Fast	50	50	52.11
Slow	53.23	58.75	53.99
Very Slow	50	50	51.03

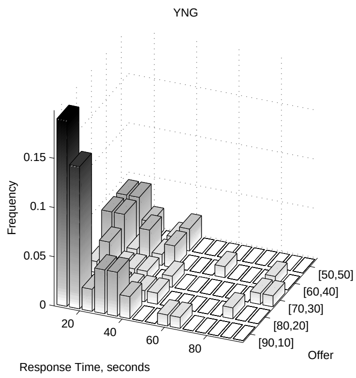
Table 2: Statistics for the distribution of expected payoffs for proposers in the UG, by *RT* category.



(a) Empirical PDF of RT for the UG and the YNG.

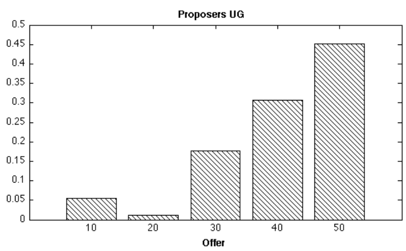


(b) UG: Joint histogram of RT and Offer.

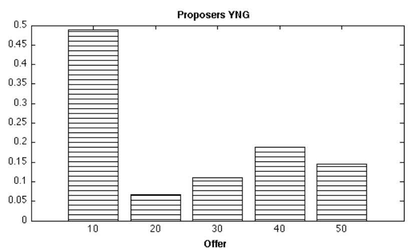


(c) YNG: Joint histogram of RT and Offer.

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.

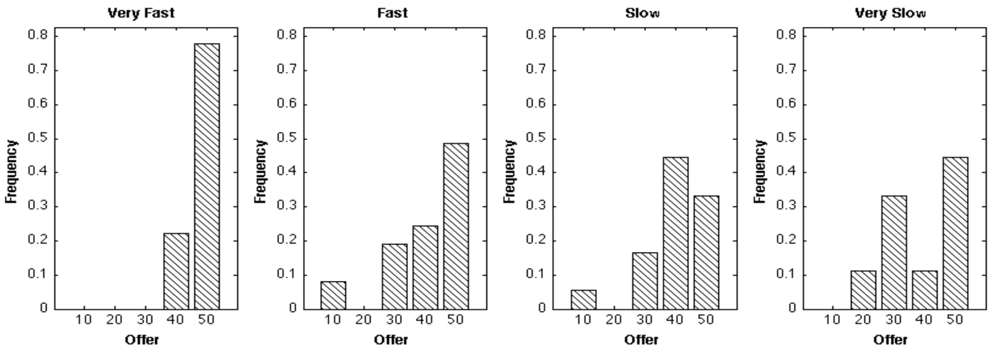


(a) Proposer offers in the UG.

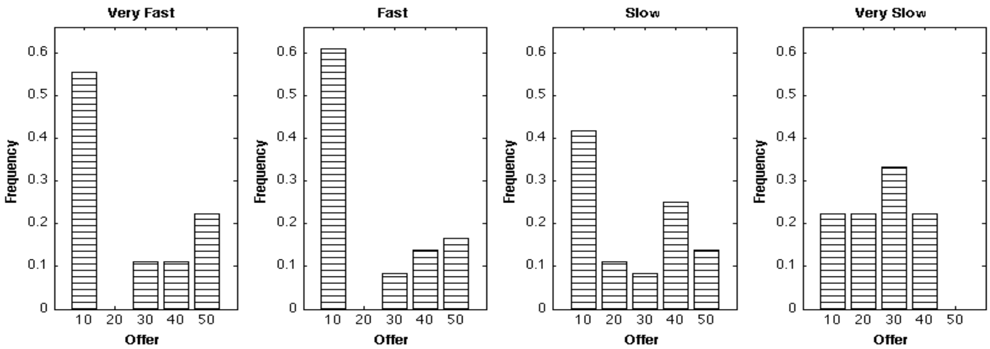


(b) Proposer offers in the YNG.

Figure 2: Histograms of offers $[100 - x, 100]$, where $x \in \{10, 20, 30, 40, 50\}$.

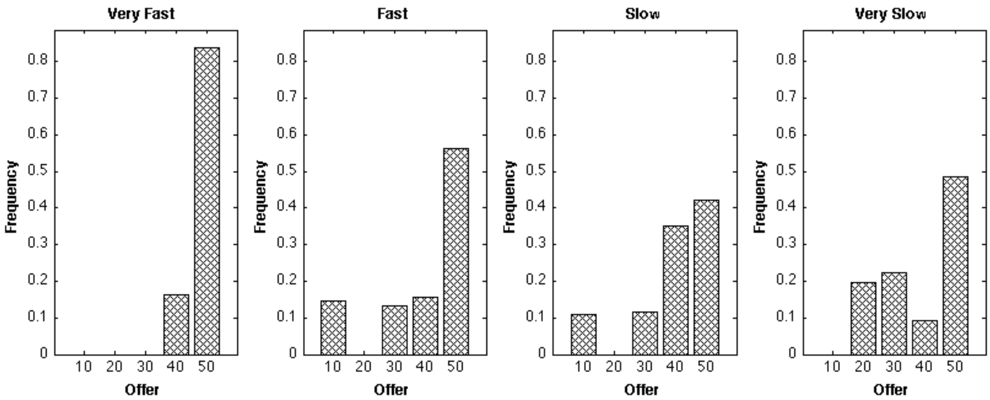


(a) Histograms of offers in each RT category for proposers in the UG.

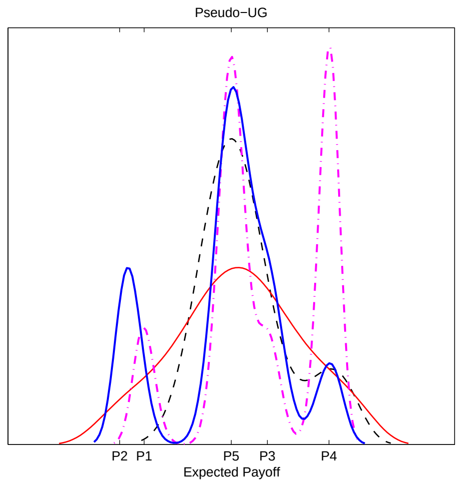


(b) Histograms of offers in each RT category for proposers in the YNG.

Figure 3: 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.



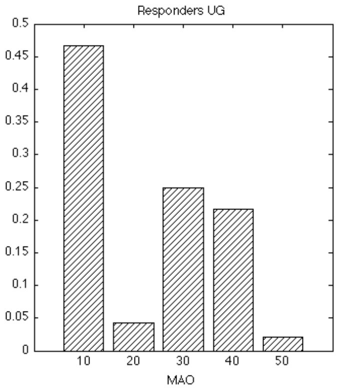
(a) Histogram of offers for different *RT* categories, *Pseudo-UG*.



	Mean
Very Fast	51.43
Fast	50.65
Slow	52.61
Very Slow	49.72

(b) Empirical cdf of expected payoff given responder MAO distribution.

Figure 4: *Pseudo-UG* sample. Distributions of offers and expected payoff by *RT* categories.



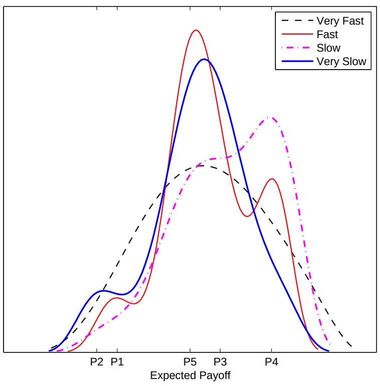
(a) Histogram of responder MAOs.

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

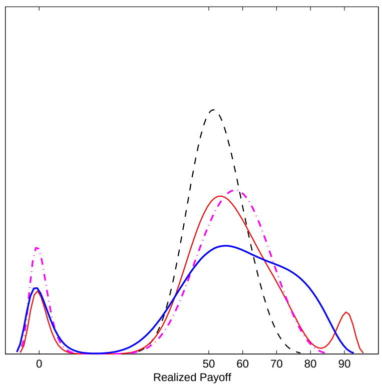
$x \rightarrow$	10	20	30	40	50
G_x	*0.4688*	*0.5104*	*0.7604*	***0.9792***	*1*
P_x	42.19	40.83	53.23	**58.75**	50
SE_x	44.91	39.99	29.88	**8.57**	0

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

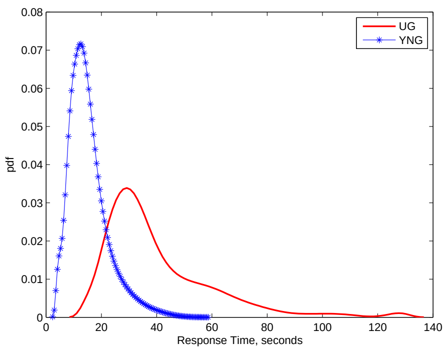


(a) Empirical cdf of expected payoff for each RT category, implied by offers made in the UG.

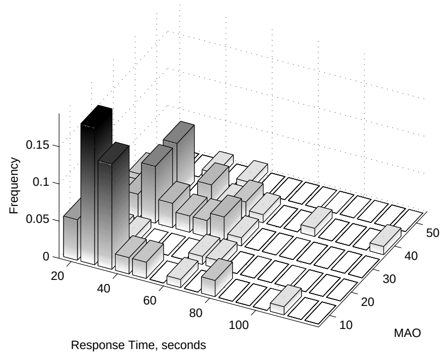


(b) Empirical cdf of realized payoffs for each RT category.

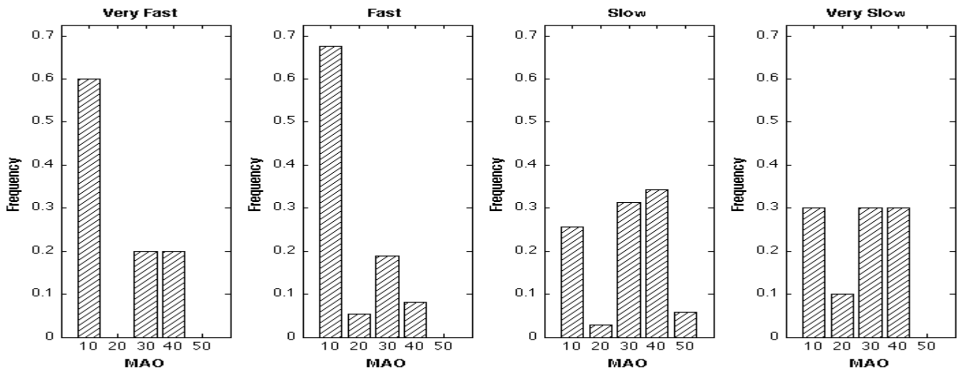
Figure 6: Expected and realized payoffs of proposers in the UG.



(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 7: Data on choices and RT for responders.

Appendices

A 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

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.

B Game Presentations

Figures A.1 to A.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*.

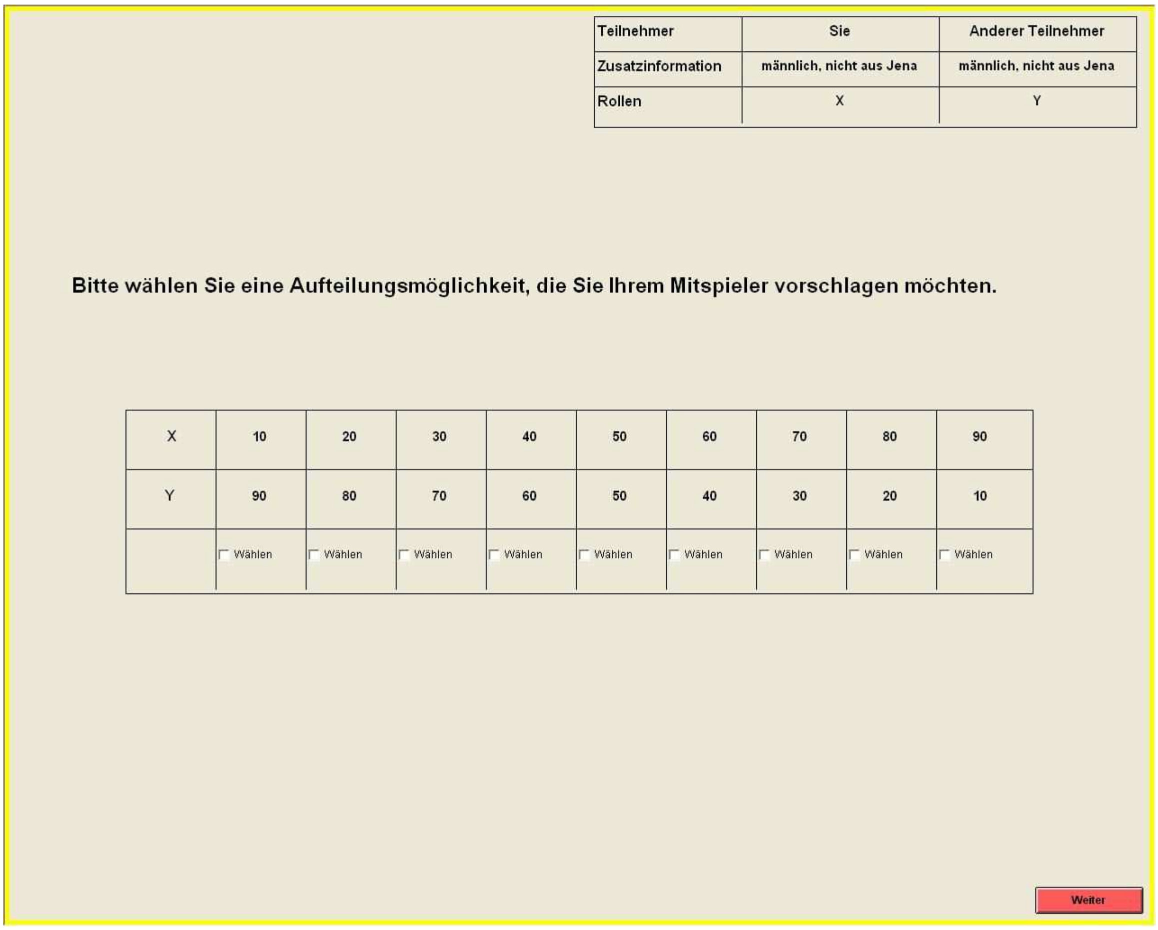


Figure A.1: Screen where proposers in either the UG or the YNG were presented their available options and made their choice.

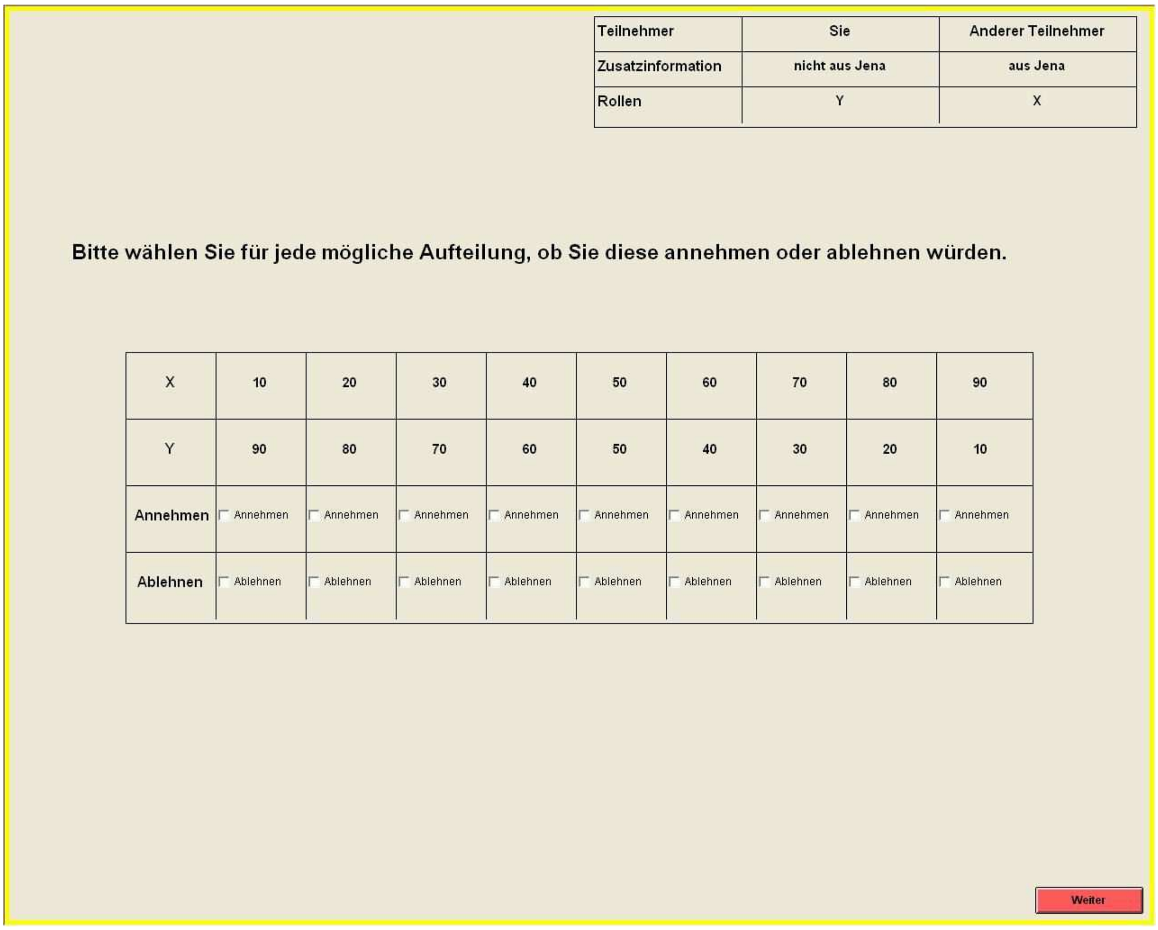


Figure A.2: Screen where responders in the UG were presented their available options and made their choice.

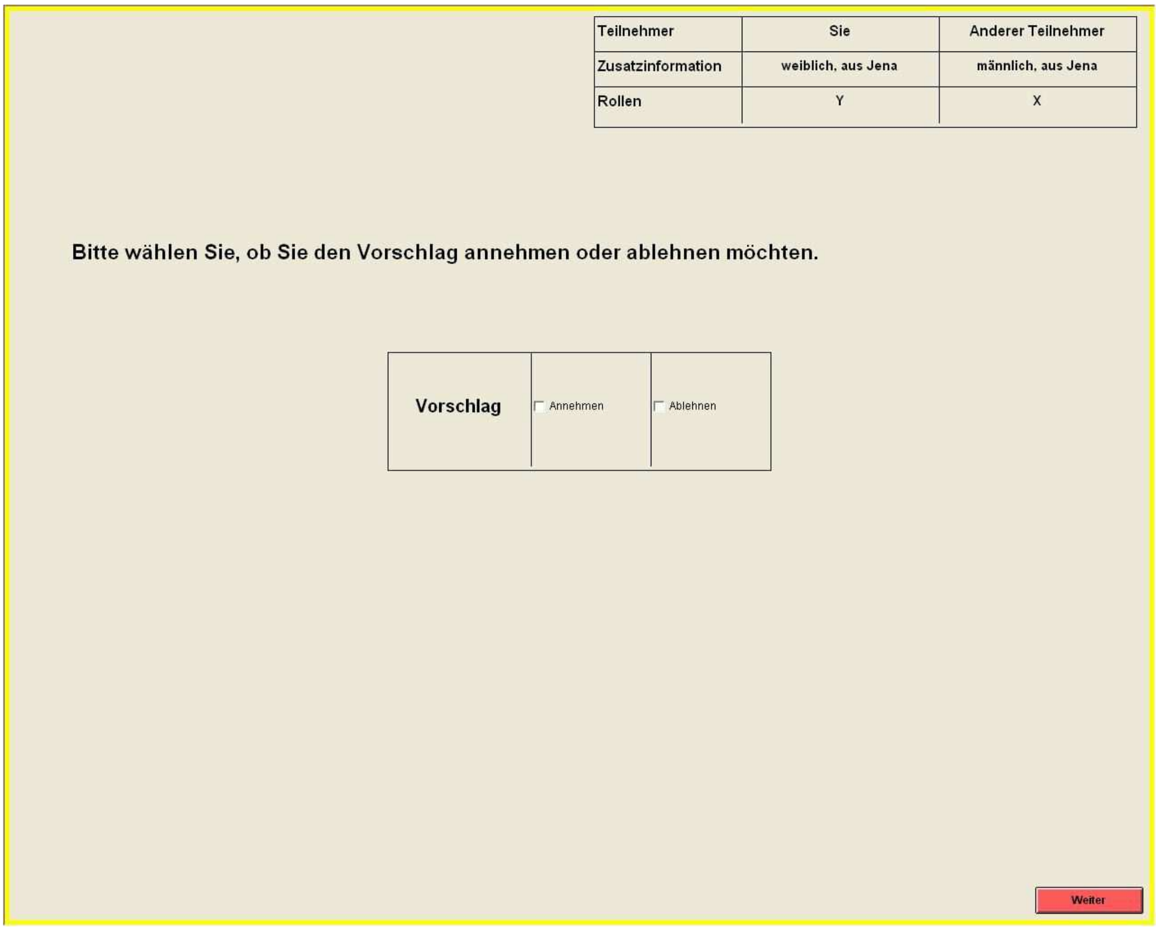


Figure A.3: Screen where responders in the YNG were presented their available options and made their choice.