INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Frustration and Anger in the Ultimatum Game: An Experiment

Chiara Aina, Pierpaolo Battigalli, Astrid Gamba

Working Paper n. 621

This Version: March, 2020

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

Frustration and Anger in the Ultimatum Game: An Experiment

Chiara Aina* Pierpaolo Battigalli† Astrid Gamba‡

Abstract

In social dilemmas, choices may depend on belief-dependent motivations enhancing the credibility of promises or threats at odds with personal gain maximization. We address this issue theoretically and experimentally in the context of the Ultimatum Minigame, assuming that the choice of accepting or rejecting a greedy proposal is affected by a combination of frustration, due to unfulfilled expectations, and inequity aversion. We increase the responder's payoff from the default allocation (the proposer's outside option) with the purpose of increasing the responder's frustration due to the greedy proposal, and thus his willingness to reject it. In addition, we manipulate the method of play, with the purpose of switching on (direct response method) and off (strategy method) the responder's experience of anger. Our behavioral predictions across and within treatments are derived from the theoretical model complemented by explicit auxiliary assumptions, without relying on equilibrium analysis.

Keywords: Experiments, psychological games, ultimatum minigame, frustration, anger, non-equilibrium analysis.
JEL classification: C72, C91, D03.

*University of Zurich. Blümlisalpstrasse 10, 8006 Zürich, Switzerland; e-mail: chiara.aina@econ.uzh.ch.

†*Corresponding author*. Bocconi University and IGIER. Via Guglielmo Röntgen 1, 20136 Milano, Italy; e-mail: pierpaolo.battigalli@unibocconi.it.

‡Università degli Studi dell'Insubria. Via Monte Generoso 71, 21100 Varese, Italy; e-mail: astrid.gamba@uninsubria.it.

1 Introduction

In this paper we test experimentally the theory of frustration and anger by Battigalli, Dufwenberg, and Smith (2019)—henceforth BDS—in the strategic context of the Ultimatum Minigame. In this social dilemma, a proposer can decide to make a fair offer that is automatically accepted, the "default allocation", or a greedy offer that the responder can either accept or reject. Experimental results about behavior in this and in similar negotiation scenarios (e.g., Ultimatum Game) systematically deviate from the predictions of traditional economic theory, reporting a positive frequency of both default allocations and rejections of greedy offers.[1] We ask whether rejections of the greedy offer can be explained by belief-dependent preferences in line with BDS theory of anger from blaming behavior. Upon observing the greedy offer, the responder may experience frustration, which is measured by the gap between his initially expected payoff and the payoff obtained accepting the greedy offer. Frustration triggers anger that may result in a rejection, depending on the responder's sensitivity to this emotion.

In the experiment, subjects play one-shot and anonymously an Ultimatum Minigame (henceforth UMG). Our experimental manipulation is twofold, affecting both the payoff structure and the method of play in a 2×2 design. We manipulate the responder's monetary payoff from the default allocation to increase his initial expectations and thus his frustration in case of a greedy offer. In addition, we manipulate the method of play, i.e., direct response versus strategy method, to determine whether responders actually observe the greedy offer before making their choice and can thus experience frustration. Indeed, when the game is played with the strategy method, responders have to decide whether to accept or reject the greedy offer without observing whether it was actually made. According to BDS theory, anger can only be triggered by the observation of an event that was initially unexpected. Thus, anger should not affect the responder's choice with the strategy method.

Subjects are assigned to one of four treatments, characterized by the method of play and the responder's (Player B) monetary payoff from the default allocation, which is either m_{b}^1 (baseline payoff treatment) or $m_{\mathrm{b}}^2 > m_{\mathrm{b}}^1$ (high payoff treatment). In every treatment, prior to choices in the experimental game, they face a prediction task through which we elicit beliefs about the behavior of subjects playing in the opponent's role. Since frustration is rooted in unfulfilled expectations, our behavioral

[1] A large number of studies, conducted with different incentives and in different countries, found a robust pattern in the Ultimatum Game: the majority of proposers offer 40% to 50% of the total sum (with the modal offer being an even split), while responders reject offers below 20% (Güth *et al.* 1982, Thaler 1988, Güth and Tietz 1990, Bolton and Zwick 1995, Hoffman *et al.* 1996, Slonim and Roth 1998, Cameron 1999, Carpenter *et al.* 2005, Andersen *et al.* 2011).

predictions across payoff treatments in the experimental game crucially depend on the responder's initial belief about the default allocation, which is endogenous and may vary with the treatment.

Given that only the responder can be affected by frustration and anger, our theoretical analysis mostly focuses on the psychological motivations that drive his decision given his initial belief about the proposer's behavior. We derive predictions about responders assuming that they have belief-dependent preferences, given by a combination of anger and inequity aversion. Predictions about the proposers are instead derived from the simplifying assumption that they maximize their expected monetary payoff.

Our theoretical analysis takes into account that the anonymous interaction of experimental subjects necessarily features incomplete information about co-players' relevant traits. In addition, the textbook equilibrium hypothesis that players have correct conjectures about the co-players' decision rules is unjustified in this context. Therefore, we just assume that (i) both players choose best replies given their personal traits and beliefs, (ii) proposers take this into account, and (iii) some mild across-treatment restrictions on beliefs hold. Assumption (i) is a *subjective rationality* condition. In the case of responders, it involves addressing a trade-off between increasing one's own payoff and decreasing the payoff of the proposer conditional on the greedy offer. Frustration and anger make this trade-off belief-dependent: the higher the initially expected payoff, the higher the incentive to reject the greedy offer. For what regards beliefs, across-payoff treatments assumptions matter only for the direct response method, as the belief-dependent motive of anger matters only in this condition. Specifically, we assume that responders in the high payoff treatment m^2_{b} are not less optimistic than responders in the baseline treatment m^1_{b}. We also assume that the relation between the distributions of the proposer's second-order beliefs in m^1_{b} and m^2_{b} reflects the relation between the distributions of the responder's first-order beliefs in these two treatments.

Given our experimental design, the theoretical analysis delivers comparative behavioral predictions across treatments. We predict a higher rejection rate with the direct response method than with the strategy method, because the frustration motive is assumed to be relevant only in the former. Focusing on the direct response method, our auxiliary comparative assumption about the responder's beliefs implies that his initially expected payoff is higher in the high payoff treatment m^2_{b} than in the baseline treatment m^1_{b}. With this, the frustration motive implies a higher frequency of rejections in the high payoff treatment m^2_{b} than in the baseline treatment m^1_{b}. Instead, the distribution of responders' choices is expected to be constant across payoff treatments with the strategy method, where rejections can only be caused by inequity aversion.

Again focusing on the direct response method, proposers are assumed to understand that the probability of rejection is increasing in the responder's initially expected payoff. With this, our auxiliary assumption about the distributions of the proposers' second-order beliefs across payoff treatments implies that they deem rejections more likely in the high payoff treatment $m_{\rm b}^2$ than in the baseline treatment $m_{\rm b}^1$. Therefore, we predict a higher frequency of the default allocation in the high payoff treatment with the direct response method. With the strategy method, instead, proposers' beliefs about conditional rejection and their choices in the game should be constant across payoff treatments because they expect the frustration motive to be absent. For the same reason, we predict more greedy offers with the strategy than with the direct response method.

Importantly, the behavioral predictions of our model hinge on whether the responder can actually experience frustration upon observing the greedy offer, which is not the case with the strategy method. As we discuss in Section 7, by adopting a broader definition of frustration—different from the one proposed in BDS—and admitting that responders can to some extent imagine the feeling of frustration and react accordingly, we can in principle allow for a positive effect of the payoff increase on rejections with the strategy method. However, it is rather implausible that imagining oneself being frustrated could produce the same angry response as actually feeling frustrated.[2] Thus, anger should at least be significantly attenuated with the strategy method compared to the direct response method.

The simplifying assumption that proposers merely maximize their expected payoff affects the predictions about their behavior. If proposers are inequity averse, an increase in the default-allocation payoff of the responder (only) may make them more prone to make the greedy offer. Since we measure proposers' beliefs, such counterbalancing effect may be revealed by the data. We discuss in Section 7 the implications of a generalized model that allows proposers to be inequity averse. Note, however, that the aforementioned predictions about responders are unaffected, because they do not rely on strategic reasoning of the responder about the proposer.

In line with the theory of frustration and anger, we find that, with the direct response method, responders' tendency to reject is positively correlated with the elicited initial expectation of the default allocation, and that the rejection rate is higher with the direct response method than with the strategy method. Also, the payoff treatment does not significantly affect behavior with the strategy method; yet, contrary to our prediction, neither it does with the direct response method.

[2] In the psychological literature, it is well-documented that the more vivid is the mental representation of an event, the greater is the emotional response (e.g., Miller *et al.* 1987).

2 Related literature

Besides reporting the closest literature on frustration and anger, we consider (1) the experimental literature focusing on the role of initial expectations in games with costly punishment actions, and (2) the experimental literature (both in economics and in psychology) that studies the role of emotions in Ultimatum-like games. We then explain how we link these two strands of the literature and how we innovate on them.

To the best of our knowledge, Persson (2018) is the only published article that tests a version of BDS theory of frustration and anger.[3] In the context of a game where unexpectedly low payoffs are determined by the simultaneous choices of chance and of another player, he finds that, while unfulfilled expectations about material payoffs generate negative emotions (as in BDS theory), these emotions do not affect subjects' punishing behavior. This result does not support some specifications of BDS theory, either simple anger, or the "could-have-been" version of anger from blaming behavior, both of which hinge on how exactly frustrated players blame their co-players. By testing BDS theory in the context of the UMG, we instead can focus on the core of BDS theory, without addressing specific versions of blaming. Another difference between our work and Persson (2018) is that his theoretical model assumes complete information about the punisher's anger sensitivity and predicts behavior using equilibrium analysis.

Our paper relates to the experimental literature on second-party punishment. In particular, a strand of this literature focuses on the role of expectations as a crucial determinant of costly punishment actions in laboratory experiments.[4] A series of papers investigating the Power-To-Take game (Bosman and van Winden 2002, Bosman *et al.* 2005, Reuben and van Winden 2008) find that expectations on the take rate affect the decision to retaliate. Some experimental studies on the Ultimatum Game show that informing responders, prior to their decision, about the average amount offered in their session (Bohnet and Zeckhauser 2004), or in previous sessions (Sanfey 2009), affects the probability of offer-specific rejections. By eliciting responders' beliefs, Azar *et al.* (2015) find a similar correlation between rejections and expected offers. Results from neuroeconomics studies also confirm the crucial role of expectations: players are more likely to reject offers that deviate

[3]There are also two recent unpublished studies by Dufwenberg *et al.* (2019a,b) that explore the role of communication—respectively promises in hold-up games and threats in deterrence games—when subjects are motivated by frustration and anger as in BDS theory.

[4]Empirical studies show similar findings. In particular, it has been found that there is a correlation between unexpected losses suffered by local teams and the number of reports of domestic violence (Card and Dahl 2011) and violent crime (Munyo and Rossi 2013).

from their expectations and the magnitude of the deviations correlate with brain areas related to negative emotions (Chang and Sanfey 2013, Xiang *et al.* 2013).

Another strand of the literature on second-party punishment focuses on the role of subjects' preferences. In economics, most of the experimental literature associates costly punishment to inequity aversion and fairness considerations. This is due to the fact that the observed frequency of rejections usually increases as the amount offered to responders decreases. Recently, a stream of literature has argued that motivations other than inequity aversion may explain these findings. The experimental study on the Ultimatum Game by Xiao and Houser (2005) shows that the possibility of expressing negative emotions (anger and disapproval) directly to proposers in reaction to greedy proposals decreases the frequency of rejections. In a similar experiment, van Kleef *et al.* (2004) show that proposers become more generous towards responders who express anger-like emotions. A recent study by van Leeuwen *et al.* (2017) confirms this result and provides experimental support to the hypothesis that anger serves as a credible commitment device (cf. Frank 1987, 1988).[5] First, responders who reject low offers display angry facial expressions (measured by a face-reader software). Second, external observers are able to detect better than a random device subjects who then reject greedy offers, on the basis of their pictures only, taken prior to receiving any information about the game.

Finally, the experiments of Grimm and Mengel (2011) and Oechssler *et al.* (2015) show that time may mitigate negative emotions arising with greedy offers: delaying the responder's decision increases the acceptance rate for low offers. Gneezy and Imas (2014) similarly find that waiting reduces anger. Conversely, Sutter *et al.* (2003) show that forcing the responder to answer within a short time increases rejection rates. Also, the rejection rate in the Ultimatum Game is higher when the allocation is decided by a person with respect to the case where an algorithm splits the money randomly between the two players (Blount 1995). All these studies suggest not only that negative emotions like anger play a role in Ultimatum-like games, but also that traditional models of social preferences developed within the standard consequentialist framework of decision theory cannot fully account for rejections, which—according to the theory—should only depend on the alternative allocations implementable by the responder for any given offer. With this, the responder's initial expectations, the source of the offer, unchosen offers, and delay should all be irrelevant. They are, instead, relevant in BDS theory and (for the role

[5] A similar effect of the strategic use of anger is found by Gneezy and Imas (2014) in an experiment involving two-player zero-sum games. Subjects seem to anticipate the emotional reaction of their co-players and exploit the strategic opportunity to anger their opponents in scenarios where this can be advantageous.

of delay) its extension to multi-period games (see Battigalli *et al.* 2019a, Section 2.3).

A strand of the psychological literature also provides empirical support for anger as the underlying mechanism for costly punishment. Pillutla and Murnighan (1996), relying on self-reports of emotions, find that low offers are often rejected and interpret rejections as angry reactions to perceived unfairness. Importantly, rejections are better predicted by reported anger than perceived unfairness. Seip *et al.* (2014) find that, both in a public good game and in a trust game, the lower the level of cooperation by co-players, the more anger is evoked and the harsher the punishment. In addition, in the trust game, increased anger (manipulated through an unrelated task) produces more severe punishments. Notice that, similarly to BDS theory, Seip *et al.* (2014) interpretation is also derived by the appraisal theory and anger is interpreted as the action tendency of a negative emotion. Yet, in Seip *et al.* (2014) anger derives from the appraisal of co-player's unfairness, and not from the appraisal of an unexpected event that blocks one's goals as in BDS theory.

Several neuroeconomics experiments employ the use of functional magnetic resonance imaging (fMRI) to study the neural activation of cognitive emotional processes in the Ultimatum Game. These studies reinforce the idea that rejections of lower offers are linked to higher activity of brain regions related to emotional decision making (Gospic *et al.* 2011, Gilam *et al.* 2018) and in particular to negative emotions (Sanfey *et al.* 2003, Gabay *et al.* 2014).[6]

All this evidence taken together suggests that there is a relation between expectations and rejections and that negative emotions like anger play a role in Ultimatum-like games. Yet, none of the papers cited above provides a theoretical interpretation that can jointly account for the role of expectations and anger in these social dilemmas, as we do in this paper. Building on previous evidence, we interpret rejections through the lenses of the non-consequentialist BDS theory that links expectations, anger-like emotions and costly punishment actions.[7] Importantly, consequentialist theories of choice do not predict any difference in the responders' behavior across different methods of play, i.e., the direct response and the strategy method. In the UMG, the only contingency in which the responder's conditional choice affects the consequences of interaction is when the proposer chooses the greedy offer. Therefore,

[6]Similar studies (van Wout *et al.* 2006, Civai *et al.* 2010) analyze skin conductance, a measure of emotional arousal, of responders who are asked to accept or reject a series of offers. Results show a stronger emotional activation when subjects face a greedy offer and its magnitude seems to be a predicting factor for rejections.

[7]Intention-based reciprocity is an alternative departure from the consequentialist interpretations of rejections (Charness and Rabin 2002 and Falk *et al.* 2003). In Section 7, we will discuss the implications of reciprocity in our strategic setting.

covert commitment to a conditional response (strategy method) should be equivalent to planning the same response and personally executing it upon observing the greedy offer (direct response method). The effect of the method of play has been debated in experimental economics for a long time due to the relatively higher potential of the direct response method in triggering emotional responses (Brandts and Charness 2011). Experimental studies on the Ultimatum Game have shown mixed evidence of behavioral differences across methods of play.[8] We contribute to this literature as we derive behavioral predictions across methods of play from BDS belief-dependence assumption, which is key to interpret a higher emotional response with the direct response method.

Finally, we deviate from most of the related literature in the way we derive our across-treatments behavioral predictions. We neither use equilibrium analysis, nor merely rely on correlations between behavior and elicited beliefs. As in Battigalli *et al.* (2013), we derive comparative predictions from the assumption that the second mover is subjectively rational (given his psychological preferences) and the first mover takes this into account. Unlike Battigalli *et al.* (2013), our auxiliary assumptions compare distributions of beliefs in different treatments.

3 Theory of frustration and anger in the UMG

We study the implications of the theory of frustration and anger in the strategic context of the UMG (Binmore *et al.* 1995), a simple binary-choice version of the Ultimatum Game (Güth *et al.* 1982). The game form with material payoffs is represented in Figure 1, where $h > m_i > \ell > 0$, $i = \mathrm{a}, \mathrm{b}$.[9] In this highly stylized social dilemma, the first mover can either implement a default allocation (d), whereby both players receive a similar amount of money, or make a "greedy" offer (g). The second mover can either accept or reject the greedy offer.

Models of social preferences are usually invoked to explain rejections in the Ultimatum Game. The most prominent of them is the inequity aversion model by Fehr and Schmidt (1999). Inequity averse preferences predict a positive rejection

[8] For example, Güth *et al.* (2001) find that varying the payoff distribution of the outside option affects proposers' behavior only with the direct response method, while—according to Brandts and Charness (2011)— their data do not reveal a significant effect of the method of play on rejections. Oxoby and McLeish (2004) find that, while there is no difference in the mean acceptance rate across methods of play, rejections of small offers are less frequent with the strategy method. The experiment by Brosig *et al.* (2003) shows that second movers of a leader-follower game engage more frequently in costly punishment of first movers with the direct response than with the strategy method (when the cost of punishment is not too high).

[9] In the classic version, the two players are given an amount of money to split, thus it has to be the case that $m_\mathrm{a} + m_\mathrm{b} = h + \ell$.

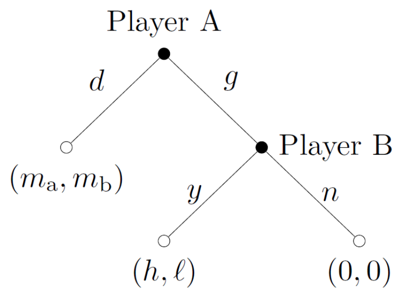


Figure 1: The Ultimatum Minigame

rate in the UMG: an inequity averse player B is willing to reject the greedy offer if he prefers the equal zero-payoff allocation over the unequal one (h, ℓ). Notice that in the inequity aversion model—as in other models of distributional preferences—player B's utility is purely outcome-based. Hence, only the comparison between the two outcomes available after the greedy offer matters. This, in turn, implies that covertly committing to a conditional response (strategy method) is equivalent to choosing the preferred allocation upon observing the greedy offer (direct response method).

Instead, according to BDS theory, player B's payoff from the default allocation can influence his decision with the direct response method as it determines his initially expected payoff and thus the extent of his frustration after the greedy offer (see Battigalli *et al.* 2019b, Example 3). Since frustration can be experienced only upon observing the greedy offer, it is muted by the strategy method.[10] These features of the theory motivate the two experimental manipulations that we conduct. As we will explain in Section 4.1, we use both methods of play (P) and, for each one of them (P=D and P=S), we change the payoff of the responder from the default allocation (m_{b}).

Since frustration and anger can only affect the responder's behavior, in this section we study only player B's decision and derive the behavior strategy associated to his psychological utility. We embed the psychological component of frustration and anger into a model that encompasses also the inequity aversion motive. Thus, we assume that player B's choice is explained by the maximization of a psychological "decision utility" function,[11] which is affected both by anger (Battigalli *et al.* 2019b) and inequity aversion (Fehr and Schmidt 1999).

[10]In Section 7 we provide a discussion of a generalization of BDS theory applied to the UMG that encompasses the possibility that frustration and anger affect the responder's behavior even with the strategy method.

[11]For a discussion of the difference between decision and experience utility see Battigalli *et al.* (2019a), Section 6.

We denote with β the initial first-order belief of player B about player A choosing the default allocation, i.e., $\beta = \mathbb{P}_{\mathrm{b}}(d)$, and with γ B's initial first-order belief about himself accepting the greedy offer, i.e., B's planned probability of acceptance $\gamma = \mathbb{P}_{\mathrm{b}}(y|g)$. B's first-order beliefs β and γ affect his decision utility through frustration. While with the strategy method frustration is equal to zero, with the direct response method it is the gap (if positive) between player B's expected payoff at the root of the game—determined by β and γ—and the payoff that he can obtain by accepting the greedy offer, i.e., ℓ:

$$F^{\mathrm{P},m_{\mathrm{b}}}(\beta,\gamma) = \begin{cases} \max\{0, \beta m_{\mathrm{b}} + (1-\beta)\gamma\ell - \ell\} & \text{if P=D} \\ 0 & \text{if P=S.} \end{cases} \tag{1}$$

With the direct response method, player B is more frustrated (i) the more he expects the default allocation, i.e., the higher β, and (ii) the more he plans to accept the greedy offer at the beginning of the game, i.e., the higher γ.

The more frustrated B is, the more he is willing to sacrifice his monetary payoff to decrease the greedy proposer's payoff. This is represented by the maximization of the following belief-dependent decision utility function with respect to action $a_{\mathrm{b}} \in \{y, n\}$ given the greedy offer g:

$$u_{\mathrm{b}}^{\mathrm{P},m_{\mathrm{b}}}(g, a_{\mathrm{b}}; \beta, \gamma) = \pi_{\mathrm{b}}(g, a_{\mathrm{b}}) - \pi_{\mathrm{a}}(g, a_{\mathrm{b}})\,\theta F^{\mathrm{P},m_{\mathrm{b}}}(\beta,\gamma) - \delta\max\{0, \pi_{\mathrm{a}}(g, a_{\mathrm{b}}) - \pi_{\mathrm{b}}(g, a_{\mathrm{b}})\}, \tag{2}$$

where $\pi_i(g, a_{\mathrm{b}})$, $i = \mathrm{a}, \mathrm{b}$, is the monetary payoff of player i after the greedy offer and B's reply a_{b}, and θ and δ are B's personal traits which measure (respectively) his sensitivity to anger and his inequity-aversion. In particular, the second term in eq. (2) represents B's inclination to harm A due to frustration and anger.

Hence, considering the two methods of play separately and replacing $F^{\mathrm{P},m_{\mathrm{b}}}(\beta,\gamma)$ with the corresponding expressions from equation (1), we obtain:

$$u_{\mathrm{b}}^{\mathrm{S},m_{\mathrm{b}}}(g, a_{\mathrm{b}}; \beta, \gamma) = \begin{cases} \ell - \delta(h-\ell) & \text{if } a_{\mathrm{b}} = y \\ 0 & \text{if } a_{\mathrm{b}} = n \end{cases}$$

with the strategy method[12] and

$$u_{\mathrm{b}}^{\mathrm{D},m_{\mathrm{b}}}(g, a_{\mathrm{b}}; \beta, \gamma) = \begin{cases} \ell - h\theta\max\{0, \beta m_{\mathrm{b}} + (1-\beta)\gamma\ell - \ell\} - \delta(h-\ell) & \text{if } a_{\mathrm{b}} = y \\ 0 & \text{if } a_{\mathrm{b}} = n \end{cases}$$

[12] Barring the trivial case where B is ex ante certain of the default allocation ($\beta = 1$), maximizing utility conditional on g is equivalent to maximizing ex ante expected utility.

with the direct response method. With the strategy method, player B (conditionally) accepts the greedy offer if and only if

$$u_{\mathrm{b}}^{\mathrm{S},m_{\mathrm{b}}}(g,y;\beta,\gamma)=\ell-\delta\left(h-\ell\right)\geq 0=u_{\mathrm{b}}^{S,m_{\mathrm{b}}}(g,n;\beta,\gamma),$$

that is,

$$\delta\leq\hat{\delta}^{\mathrm{S}}:=\frac{\ell}{h-\ell}.$$

Player B's initial expectations and sensitivity to anger kick in when the game is played with the direct response method. In this case, player B accepts the greedy offer if and only if

$$u_{\mathrm{b}}^{\mathrm{D},m_{\mathrm{b}}}(g,y;\beta,\gamma)=\ell-h\theta F^{\mathrm{D},m_{\mathrm{b}}}\left(\beta,\gamma\right)-\delta(h-\ell)\geq 0=u_{\mathrm{b}}^{\mathrm{D},m_{\mathrm{b}}}(g,n;\beta,\gamma).$$

To provide a unified representation of the best reply correspondence, we express acceptance as a condition on δ also for the direct response method, but in this case the threshold depends on B's belief β, his planned probability of acceptance γ, and his sensitivity to anger θ:

$$\delta\leq\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}\left(\beta,\gamma,\theta\right):=\frac{\ell-h\theta\max\left\{0,\beta m_{\mathrm{b}}+\left(1-\beta\right)\gamma\ell-\ell\right\}}{h-\ell}. \tag{3}$$

The following proposition summarizes:

Proposition 1. *Given the method of play* P *and B's payoff from the default allocation* m_{b}*, player B accepts the greedy offer if and only if his degree of inequity aversion is low enough, i.e., if and only if*

$$\delta\leq\hat{\delta}^{\mathrm{P},m_{\mathrm{b}}}(\beta,\gamma,\theta):=\frac{\ell-h\theta F^{\mathrm{P},m_{\mathrm{b}}}\left(\beta,\gamma\right)}{h-\ell}=\begin{cases}\frac{\ell-h\theta\max\{0,\beta m_{\mathrm{b}}+(1-\beta)\gamma\ell-\ell\}}{h-\ell} & \textit{if } \mathrm{P}=\mathrm{D}\\ \frac{\ell}{h-\ell} & \textit{if } \mathrm{P}=\mathrm{S}.\end{cases}$$

The proposition shows that the acceptance condition is easier to meet with the strategy method than with the direct response method. Furthermore, changes in the default allocation payoff m_{b} can affect B's behavior only when the UMG is played with the direct response method. When m_{b} increases, other things being equal, B's expected payoff also increases and so does his frustration, making it more difficult to meet the acceptance condition 3.

When P = D and anger may affect B's behavior, his reply to the greedy offer also depends on his plan of accepting it. Specifically, B's willingness to accept after he observed the greedy offer is *decreasing* in the planned probability to accept, γ. For some values of (β,θ,δ), if B plans to accept, the resulting expected payoff is high enough to make him sufficiently frustrated by the greedy offer and willing to reject it; similarly, if B plans to reject, the resulting expected payoff is low enough to make him willing to accept. Such dependence of preferences on one's own planned behavior is typical of several models of belief-dependent motivation and gives rise to a rationality condition of a fixed-point kind: planned behavior must be consistent with the incentives determined by the plan itself, that is, it must form an *intrapersonal equilibrium*.[13] In Appendix 1 we derive B's intrapersonal equilibrium function $\Gamma^{\mathrm{P},m_{\mathrm{b}}}(\beta,\theta,\delta)$, which describes the incentive-compatible probability of acceptance as a function of B's relevant features, i.e., his personal first-order beliefs (possibly affected by the treatment) and his exogenous personal traits. The following result establishes that, with the direct response method, the rationally planned probability of accepting (respectively, rejecting) the greedy offer decreases (increases) with B's expectation of the default allocation.

Proposition 2. *The incentive-compatible probability of acceptance* $\Gamma^{\mathrm{P},m_{\mathrm{b}}}(\beta,\theta,\delta)$ *is decreasing in* β *with the direct response method (P=D), and it is independent of* β *with the strategy method (P=S).*

4 The experiment

In this section we present our experiment. First, we describe the treatments in detail. Next, we outline the experimental procedures.

4.1 Experimental design

We conduct a two-fold manipulation, informed by the analysis of player B's behavior strategy reported in the previous section.[14]

Payoff manipulation. With the purpose of increasing player B's frustration, we manipulate player B's material payoff from the default allocation (m_{b}), while keeping all the other payoffs constant. Notice that, varying m_{b}, player B's initial beliefs may also change—i.e., beliefs are endogenous to the treatment. In particular, player B's initially expected payoff increases if the subjective probability β initially assigned to the default allocation is at least as high in the high payoff treatment than in the baseline payoff treatment. According to the theoretical model, B's

[13]The intrapersonal equilibrium approach is standard in the analysis of sophisticated agents with dynamically inconsistent preferences. See, for example, the models of Caplin and Leahy (2001), Koszegi and Rabin (2009) and the methodological contributions by Battigalli and Dufwenberg (2009) and Battigalli *et al.* (2019a). Note that this is only an *intrapersonal* equilibrium because belief β is not assumed to be correct.

[14]We preregistered the experimental design, the hypotheses and the planned analysis on AsPredicted, available at the following link: `http://aspredicted.org/blind.php?x=ij3fg3`.

incentive to accept or reject is independent of A's payoff from the default allocation (m_a). Therefore, we keep it constant across treatments.[15] Also, we do not alter A's and B's payoff from accepting the greedy offer (h and ℓ) in order to keep constant the inequity-aversion effect. With this, in the baseline payoff treatment B-subjects obtain a payoff of $m_b^1 = 8$ EUR from the default allocation, in the high payoff treatment they obtain $m_b^2 = 11$ EUR. Figure 2 illustrates the UMG game form for m_b^1 and m_b^2.

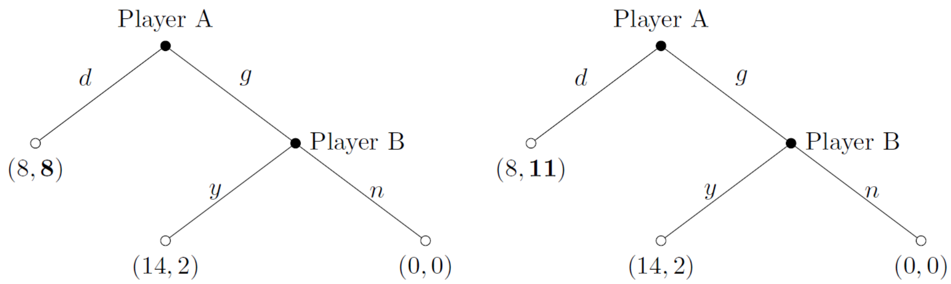


Figure 2: The UMG game form in payoff treatment m_b^1 (left) and m_b^2 (right)

Method of play manipulation. With the purpose of switching on and off player B's experience of frustration, we manipulate the method of play, i.e., whether the UMG is played with the direct response method (P = D) or the strategy method (P = S). As discussed in Section 3, when P = S player B cannot experience frustration and anger at the time of choice. Thus, B's willingness to reject the greedy offer with the strategy method is lower than with the direct response method, and it is invariant to the default allocation payoff m_b.

Table 1 summarizes our 2×2 design, reporting our four treatments and the corresponding labels. We implement a *between-subjects* design, so that subjects play only one of the treatments D1, D2, S1, and S2.

Before subjects play the UMG, we elicit their first-order beliefs with a procedure that we will explain in detail below. In our context, eliciting beliefs is particularly important for two reasons. The first and most important is that we can verify whether B-subjects' different expectations of the default allocation in treatments

[15] Falk *et al.* (2003) investigate various versions of the UMG; in particular, they compare the rejection rate in a traditional UMG, where the default allocation is (5,5) and the offer is (8,2) with an UMG where the default allocation is (2,8) and the offer is (8,2). In the latter treatment, where the default option is very unfair to the proposer, they find that the rejection rate drops from 44% to 27% and the choice of the offer (8,2) grows from 31% to 73%. We instead keep the proposer's payoff constant in the default allocation and slightly increase the responder's payoff in order to minimize the effect that varying the distribution of payoffs in the default allocation may have both on offers and rejections. Indeed, our objective is to isolate anger as the driving force of rejections and—if anticipated—of making less unfair offer.

Table 1: Experimental design

		Payoff treatment	
		m_{b}^1	m_{b}^2
Method of play	Direct response method	D1	D2
	Strategy method	S1	S2

D2 and D1 offset the difference in payoffs. The second is that it may enhance our understanding of subjects' strategic reasoning.

4.2 Experimental procedures

We recruited 352 participants from the subjects pool of Bocconi University students with the software SONA.[16] The experiment was programmed with the software oTree (Chen *et al.* 2016) and run in the BELSS laboratory of Bocconi University. Average pay was 8.70 EUR (including a show up fee of 4 EUR) and the experiment lasted on average 45 minutes. We ran 16 sessions, with 11 pairs of subjects per session, formed randomly and anonymously. To have a comparable number of observations for the responders, we had 3 sessions for each payoff treatment with the strategy method (33 observations per role) and 5 sessions for each payoff treatment with the direct response method (55 observations per payoff treatment for A, and 35 in D1 and 37 in D2 for B, respectively). Instructions were read aloud and we made sure subjects understood the rules by asking control questions.

We elicited first-order beliefs of both A and B-subjects before choices with a very simple procedure.[17] Since in every session there were 11 pairs of subjects, we asked B-subjects to guess how many A-subjects of the other pairs in the session would choose the default allocation.[18] Similarly, we asked A-subjects playing with the strategy method to guess how many B-subjects of the other pairs would choose to accept the greedy offer. Instead, we asked A-subjects playing with the direct response method to guess which percentage of B-subjects of the other pairs, among

[16]The subject pool includes undergraduate students from different disciplines (Economics, Business, Law and Political Science).

[17]We elicited beliefs before, and not after, choices as BDS theory relates player B's choice to his *initial* expectations. The evidence found by Nyarko and Shotter (2002), Costa-Gomes and Weizsäcker (2008), Ivanov (2011) and Bauer and Wolff (2018) shows that the impact of ex ante belief elicitation on subjects' behavior is negligible.

[18]One could in principle suspect that beliefs about the co-player differ from beliefs about subjects of other pairs, due to the well-documented phenomenon of over-optimism about own future outcomes or wishful thinking (e.g., Weinstein 1989, Bar-Hillel and Budescu 1995, Camerer and Lovallo 1999). This issue is addressed by Bauer and Wolff (2018), who find no evidence for the wishful thinking bias and, importantly, find no significant difference between beliefs about the co-player and about a random other subject, as can be seen in Figure 5, page 23.

those who received the greedy offer, would decide to accept it.[19] Notice that asking subjects to make a guess about the behavior of subjects' playing in the opponent's role in the *other* pairs has the purpose of avoiding additional (and confounding) emotional responses of B-subjects with the direct response method. Had they had to make guesses also on their co-player's behavior, upon observation of the greedy offer, they might have been disappointed and aggrieved by the realization of the wrong guess in the beliefs task.

All subjects were paid 10 EUR in case of a correct guess and 2 EUR in case of a wrong guess. With the purpose of avoiding hedging, subjects were either paid for the beliefs elicitation task or for the UMG play (with the same probability). Subjects in the same pair were paid for the same task in order to provide them with appropriate incentives within the UMG play.

After the experiment, we asked B-subjects to fill the State-Trait Anger Expression Inventory-2 questionnaire, namely STAXI-II (Spielberger 1999), that delivers both a measure of sensitivity to anger as a stable trait and a measure of the anger state. While B-subjects were filling the STAXI-II, A-subjects were asked to fill the Aquino questionnaire about morality (Aquino and Reed 2002) and the GASP test of guilt aversion (Cohen *et al.* 2011).[20] The STAXI-II was important to obtain a measure of B-subjects' sensitivity to anger and understand the motivation behind their behavior, while the questionnaires administered to A-subjects had the main purpose to make the duration of the experiment symmetric across A and B-subjects. Yet, both questionnaires measure personal traits that potentially affect A-subjects' behavior and a significant correlation between these measures and A-subjects' choices would refute the selfishness assumption.

In the next section we derive our behavioral predictions given the analysis of Section 3 and the experimental design described above.

5 Behavioral predictions

In Section 3 we studied the determinants of player B's decision. The goal of this section is to obtain behavioral predictions, with a focus on across-treatment comparisons. In compliance with the standard language of experimental economics,

[19] Notice that, with the direct response method, it may occur that no A-subject makes the greedy offer so that no B-subject has to take a decision, while with the strategy method B-subjects decide independently of their co-player's move.

[20] The Aquino Moral Identity Scale measures the degree to which people's self-concepts center on moral traits—i.e., how much people consider their morality to be central to their identities. The GASP test measures individual differences in the propensity to experience guilt and shame across a range of personal transgressions.

we state our predictions as "experimental hypotheses", but we stress that most of them are results derived from the analysis of Section 3 and from assumptions about players' rationality and beliefs. Two of the assumptions about beliefs are auxiliary hypotheses. All the experimental hypotheses representing results and not directly implied by the theoretical analysis of Section 3 are rigorously derived from a formal model in Appendix 2. Here we provide a sketch of the analysis with intuitive explanations.

Most economic models rely on the assumption that players' beliefs are coordinated on an equilibrium, hence that they are correct.[21] Such equilibrium assumption is sometimes a useful theoretical shortcut, but it is justified only under special circumstances, that is, either when agents play the same game recurrently and it can be shown that learning eventually leads to correct beliefs,[22] or when strategic reasoning based on whatever is common knowledge about the game yields a unique prediction. Neither of these conditions hold in our experiment: agents have no way to learn; furthermore, agents' non-selfish or belief-dependent preferences are not common knowledge, which implies that strategic reasoning cannot yield unique predictions (see the discussion and analysis in Battigalli *et al.* 2019a, and Attanasi *et al.* 2013). Therefore, our predictions do not rely on equilibrium analysis. We perform, instead, something like two steps of elimination of non-best replies, taking into account the leader-follower structure of the game form and players' incomplete information. Specifically, we assume that players are subjectively rational and, in order to derive predictions about proposers' behavior, we assume that they are certain of responders' subjective rationality.[23] For what regards beliefs, our auxiliary assumptions matter only for the direct response method, since with the strategy method preferences are not belief-dependent. Specifically, we make comparative assumptions about the distribution of beliefs across payoff treatments (D1 *vs* D2). We emphasize that we do not ascribe such comparative assumptions to the agents, i.e., we do not interpret them as coming from across-treatment reasoning, as such interpretation is prevented by our between-subjects setting.

We focus on a setting that mimics the random matching structure typical in laboratory experiments, as if experimental subjects are the players of a population game implemented by the experimenter in which subjects are randomly matched

[21] Subjective Bayes-Nash equilibrium is only a partial exception: beliefs about co-players' (endogenous) decision functions are assumed to be correct, but beliefs about their (exogenous) types are subjective, hence potentially incorrect. See Attanasi *et al.* (2016) and Battigalli *et al.* (2019a) for an analysis and discussion of subjective Bayesian equilibrium in psychological games.

[22] See, e.g., Battigalli *et al.* (1992), and Fudenberg and Levine (1998).

[23] Subjective rationality refers to an elaboration of the standard notion of rationality with subjective expected utility maximization. Player B's subjective utility coincides with his psychological utility given the greedy offer defined in condition 2 of Section 3.

with one another. We analyze the game as an interaction between two individuals, player A and player B, that are drawn at random from a large population of subjects. We first derive predictions about B-subjects' behavior and then about A-subjects' behavior.

5.1 Behavioral predictions about responders

Our behavioral predictions about responders (B-subjects) follow from the notion of subjective rationality developed in Section 3 whereby, conditional on the greedy offer, B solves a trade-off between increasing his (monetary) payoff and decreasing A's payoff, given his initial expectations and personal traits. Our first two experimental hypotheses about responders follow directly from the core of BDS theory. Proposition 2 implies the following:

Hypothesis 1.1. *With the direct response method, and not with the strategy method, the higher the initial expectation of the default allocation, the lower the probability of accepting the greedy offer.*

Next, we make comparisons about B's behavior across methods of play, for fixed payoff treatment. Both with the direct response and with the strategy method the inequity aversion component is present and may justify a rejection. However, with the direct response method, anger caused by the greedy offer can decrease even further B's willingness to accept (see Proposition 1).

Hypothesis 1.2. *The frequency of acceptance of the greedy offer is lower with the direct response than with the strategy method.*

To obtain comparative predictions on B's behavior across payoff treatments with the direct response method, on top of assuming B's subjective rationality, we make an across-payoff treatment assumption on B's beliefs: player B tends to assign a probability to the default allocation in D2 at least as high as in D1. This auxiliary assumption is testable, as it gives rise to the following experimental hypothesis.

Hypothesis 1.3. *B-subjects' average expectation of the default allocation in D2 is not lower than in D1.*

The foregoing comparative assumption about beliefs and the payoff manipulation imply that the initially expected payoff of B-subjects in D2 is higher than in D1. Therefore, we predict that B-subjects in D2 are more frustrated by the greedy offer, hence more prone to an angry rejection.[24] Instead, according to BDS theory,

[24]In Appendix 2 we show that the payoff treatment effect with the direct response method can also be obtained by substituting the intrapersonal equilibrium assumption with a plausible assumption on player B's first-order beliefs.

B-subjects cannot feel frustrated when contemplating their conditional response, which rules out payoff-treatment effects with the strategy method.

Hypothesis 1.4. *The frequency of acceptance of the greedy offer is lower in D2 than in D1, but it does not differ between S2 and S1.*

5.2 Behavioral predictions about proposers

We proceed further with the analysis of A-subjects. As anticipated in the Introduction, to make the analysis more tractable, we assume that A is "selfish" and risk neutral, i.e., that he maximizes his expected monetary payoff (in Section 7 we discuss the implications of inequity aversion by the proposer). Let $\alpha = \mathbb{P}_{\mathrm{a}}(y|g)$ denote the endogenous first-order belief of A. In each treatment, player A chooses the default allocation (d) if and only if

$$\pi_{\mathrm{a}}(d) = m_{\mathrm{a}} \geq \alpha h = \mathbb{E}_{\alpha}[\pi_{\mathrm{a}}|g],$$

i.e., if and only if

$$\alpha \leq \hat{\alpha} := \frac{m_{\mathrm{a}}}{h}.$$

Note that the threshold $\hat{\alpha}$ does not depend on the treatment, as we have purposely set A's payoffs (m_{a}) constant across treatments. This yields our first experimental hypothesis about A-subjects.[25]

Hypothesis 2.1. *The higher the expectation of acceptance, the higher the probability of making the greedy offer.*

To obtain comparative predictions on the frequency of default allocations across treatments, we assume not only that A is subjectively rational, but also that he is certain of B's subjective rationality, i.e., that B solves the trade-off (if any) between increasing his payoff and decreasing the greedy proposer's payoff as explained in Section 3. In particular, A is supposed to attach higher probability to player B accepting the greedy offer with the strategy method, where the frustration and anger motivation is muted, than with the direct response method. It follows that a rational player A is more inclined to choose the default allocation in Dk than in Sk.

Hypothesis 2.2. *The frequency of the default allocation is higher with the direct response method than with the strategy method.*

[25]Notice that evidence in favor of Hypothesis 2.1 would be compatible with the selfishness assumption that we made, but it does not fully exclude that A-subjects may be affected by some sort of social preferences (e.g., inequity aversion). Yet, we can test whether concerns for the payoff distribution drive their decision by looking at the effect of the payoff treatment on their behavior controlling for beliefs.

To compare A-subjects' behavior across payoff treatments with the direct response method (D2 *vs* D1), we make a comparative assumption concerning the distributions of A's second-order beliefs about B's expectation of the default allocation. Specifically, we assume that the distributions of A's second-order beliefs in the two payoff treatments satisfy a stochastic dominance relation in line with the relation between B's beliefs distributions: player A tends to assign a probability to B expecting the default allocation in D2 at least as high as in D1. Since we do not elicit second-order beliefs, we cannot directly test this auxiliary assumption, but we can at least test its implication concerning first-order beliefs given A's certainty of B's subjective rationality.

Hypothesis 2.3. *A-subjects' average expectation of acceptance is lower in D2 than in D1, but it does not differ between S2 and S1.*

Since the threshold $\hat{\alpha}$ below which A chooses the default allocation is constant across treatments, this implies that A-subjects tend to choose the default more often in D2 than in D1, while no difference is expected between S2 and S1 where only the inequity-aversion motive can make A afraid of a rejection.

Hypothesis 2.4. *The frequency of the default allocation is higher in D2 than in D1, but it does not differ between S2 and S1.*

6 Results

This section presents the experimental results. We start by providing a description of B-subjects' behavior. Next, we report results for A-subjects' behavior.

There were 176 B-subjects in total, with 66 facing the strategy method and 110 the direct response method. Only 72 out of the 110 B-subjects in the direct response method received the greedy offer and did actually make a choice. Thus, we have in total 138 observations of B-subjects' behavior.

Before examining B-subjects' behavior, we check whether our auxiliary assumption regarding their initial first-order belief about the default allocation is verified.[26] Table 2 shows the average subjective probabilities of the default allocation of B-subjects prior to choices in the experimental UMG, by payoff treatment and method of play. The table also reports the differences in means with the relative p-value of the tests. Results of both a two-sample t-test of the difference in means and of a Mann-Whitney U test (henceforth, MW) indicate that there is no significant difference in initial beliefs between D1 and D2, the two estimates being virtually identical.

[26]Throughout this section, with reference to hypothesis testing, we mean "(not) verified"as an abbreviation of "(not) consistent with the evidence".

Thus, our auxiliary assumption is verified: B-subjects' average expectation of the default allocation in D2 is not lower than in D1, as stated in Hypothesis 1.3.

Instead, a p-value of 0.06 (MW)—and of 0.04 for the t-test—indicates that there is a significant difference between beliefs in S1 and S2, with higher values in S1. However, notice that our predictions about B-subjects' behavior do not rely on auxiliary assumptions regarding their initial first-order beliefs with the strategy method.

Table 2: B-subjects' expectations of the default allocation

Mean β	m_b^1	m_b^2	Diff.	p-value (t-test)	p-value (MW)
Direct response	0.54	0.54	0.00	1.00	0.93
Strategy	0.62	0.48	0.15	0.04	0.06
Diff.	-0.09	0.06			
p-value (t-test)	0.14	0.31			
p-value (MW)	0.10	0.41			

Notes: means of B-subjects' subjective probability of the default allocation by payoff treatment and method of play; differences across treatments and p-values of a two-sided t-test of the differences in means and of a Mann-Whitney U test are displayed.

Figure 3 illustrates the share of B-subjects accepting the greedy offer by treatment. Overall, the average frequency of acceptance of the greedy offer is approximately 64%. The observed frequency of acceptance is lower with the direct response than with the strategy method (MW p-value 0.008; 138 observations) and it is similar across payoff treatments both with the direct response (MW p-value 0.65; 72 observations) and with the strategy method (MW p-value 0.57; 66 observations). This evidence seems to provide support to Hypothesis 1.2 (lower frequency of acceptance with the direct response method), but it is at odds with Hypothesis 1.4 (higher frequency of acceptance in D1 than in D2, but constant across S1 and S2).

Next, we analyze the effect of B-subjects' initial expectations on the probability of accepting the greedy offer. Table 3 reports the means of B-subjects' subjective probabilities of the default allocation conditional on accepting or rejecting the greedy offer (by method of play). In line with our predictions, there is a significant negative correlation between B-subjects' initial expectations and the probability of acceptance with the direct response method: B-subjects rejecting the greedy offer expected on average that the default allocation was 20% more likely than those who accepted it. We also find mild evidence of the same correlation with the strategy method. The latter was unexpected though, thus Hypothesis 1.1 is only partially verified.

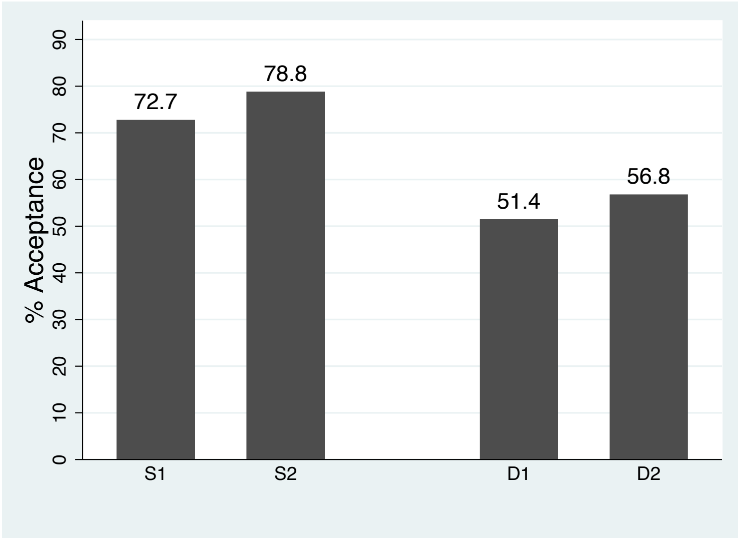


Figure 3: B-subjects' behavior by treatment

Table 3: Initial expectations and rejections

Mean β	Accept	Reject	Diff.	p-value (t-test)	p-value (MW)
Direct response	0.42	0.62	-0.20	0.00	0.00
Strategy	0.51	0.67	-0.16	0.05	0.08
Diff.	-0.09	-0.06			
p-value (t-test)	0.16	0.25			
p-value (MW)	0.18	0.17			

Notes: means of B-subjects' subjective probability of the default allocation conditional on accepting and rejecting the greedy offer, by method of play; differences across conditions and p-values of a two-sided t-test of the differences in means and of a Mann-Whitney U test are displayed.

In order to estimate the effects of our treatments on the probability of accepting the greedy offer, we use the following probit model:

$$\mathbb{P}(\mathrm{y}_i = 1|\mathrm{Direct}_i, \mathrm{Payoff\ Increase}_i, \mathrm{x}_i) = \Phi(\lambda_0 + \lambda_1 \mathrm{Direct}_i + \lambda_2 \mathrm{Payoff\ Increase}_i + \lambda_3 \mathrm{Direct}_i \times \mathrm{Payoff\ Increase}_i + \lambda_4^{'} \mathrm{x}_i), \quad (4)$$

where Φ is the standard normal cumulative distribution function, y_i, with $i = 1, ..., 138$, is a dummy that takes value 1 when subject i accepts the greedy offer, Direct_i is a dummy that takes value 1 (0) with the direct response method (strategy method), $\mathrm{Payoff\ Increase}_i$ is a dummy that takes value 1 (0) in payoff treatment m_{b}^2 (m_{b}^1) and x_i is a vector of controls that may or may not include (depending on the specification) i's beliefs (on a scale from 0 to 100) and their interaction with Direct_i, Age, Gender (dummy that takes value 1 if female) and Trait Anger (as elicited from STAXI-II).[27]

Table 4: B-subjects' behavior

	I	II	III
Direct	-0.217***	-0.246***	-0.249***
	(0.079)	(0.072)	(0.071)
Payoff Increase	0.057	0.058	0.069
	(0.079)	(0.074)	(0.074)
Direct × Payoff Increase	-0.007	0.070	0.045
	(0.158)	(0.145)	(0.143)
Belief		-0.006***	-0.006***
		(0.001)	(0.001)
Belief × Direct		-0.004	-0.003
		(0.002)	(0.002)
Trait Anger			-0.006
			(0.004)
Age			-0.042*
			(0.021)
Gender			0.064
			(0.077)
Payoff Increase if Direct = 1	0.053	0.091	0.089
	(0.117)	(0.104)	(0.104)
Payoff Increase if Direct = 0	0.061	0.021	0.044
	(0.105)	(0.101)	(0.099)
Observations	138	138	138

Notes: the dependent variable is a dummy that indicates whether i accepts the greedy offer; the table reports average marginal effects estimated from probit models; for the treatment interaction, we report the differential marginal effects of Payoff Increase when Direct changes from 0 to 1 (conditional marginal effects are reported in the lower part of the table); for the interaction of Belief with Direct, we report the differential marginal effect of Belief when Direct changes from 0 to 1; standard errors in parentheses; significance levels are: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 4 reports the average marginal effects of the treatment variables on the probability of acceptance as implied by different specifications of the probit estimation. The upper part of column I reports the average marginal effects of the two treatments and their interaction (Direct × Payoff Increase), which is computed as the difference between the marginal effects of Payoff Increase in the two methods

[27] Trait Anger measures the individual's general disposition toward anger as a stable trait, differently from State Anger that measures the individual's current emotional state. We do not use State Anger as a control since it is measured at the end of the experiment, after all choices are made and payoffs are realized, and thus may not correctly identify B-subjects' emotions at the information set following the greedy offer. Note that the Trait Anger, as expected, does not correlate with the final payoff (p-value 0.793).

of play (reported in the lower part of the table). Column II reports results from an alternative specification of equation (4) that includes Belief—i.e., i's expectation of the default allocation—and its interaction with Direct as explanatory variables. Specifically, Belief × Direct refers to the difference between the marginal effects of beliefs with the two methods of play. Column III reports the same estimates for a specification of equation (4) that includes Trait Anger, Age and Gender as controls.

In line with Hypothesis 1.2, the estimated probability of accepting the greedy offer is lower with the direct response than with the strategy method, as indicated by a negative and significant average marginal effect of Direct. Against Hypothesis 1.4, the estimates do not indicate a negative effect of the payoff increase on the probability of acceptance, the marginal effect of Payoff Increase being not significant with either method of play.

For what regards Hypothesis 1.1, B-subjects' first-order beliefs have a significant negative effect on the probability of acceptance with both methods of play. Yet, the negative impact of beliefs is amplified with the direct response method: against an increase of 10% in the expectation of the default allocation, the probability of acceptance decreases (on average) of approximately 6% points with the strategy method and of 9% with the direct response method. The negative difference of 4% points between the marginal effect of belief with the direct response and with the strategy method displays a p-value of 0.12. The negative interaction between beliefs and the direct response method is clearly illustrated in Figure 4, which shows the predicted probability of acceptance against beliefs from a probit model with beliefs, treatment dummies and their interaction as regressors. Beliefs have a negative effect on the probability of acceptance not only with the direct response method but, contrary to our predictions, also with the strategy method. The latter effect is smaller though.

Among controls, Trait Anger has a (mildly significant) negative effect (p-value: 0.13) on the probability of accepting the greedy offer. In line with the theory of frustration and anger, the higher is the sensitivity to anger, the lower is the probability of acceptance.

Let us now describe our experimental findings about A-subjects' behavior and beliefs. We have in total 176 observations for A-subjects' (66 for Strategy and 110 for Direct). The overall frequency of default allocations is approximately 36%. Figure 5 illustrates the share of A-subjects making the default allocation by treatment. The figure suggests that neither the method of play (MW: p-value 0.65; 176 observations) nor the payoff treatment with the direct response method (MW p-value 0.69; 110 observations) and with the strategy method (MW p-value 0.45; 66 observations) have a significant effect on the frequency of the default allocation.

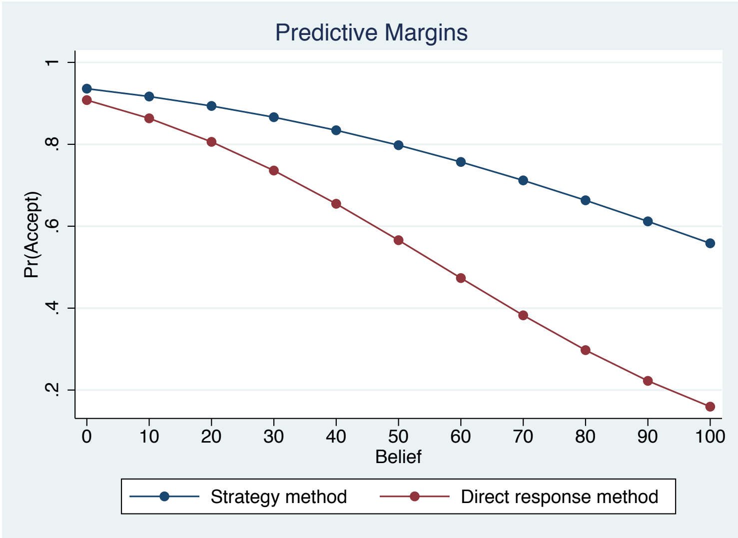


Figure 4: Predicted probability of acceptance against B's first-order beliefs

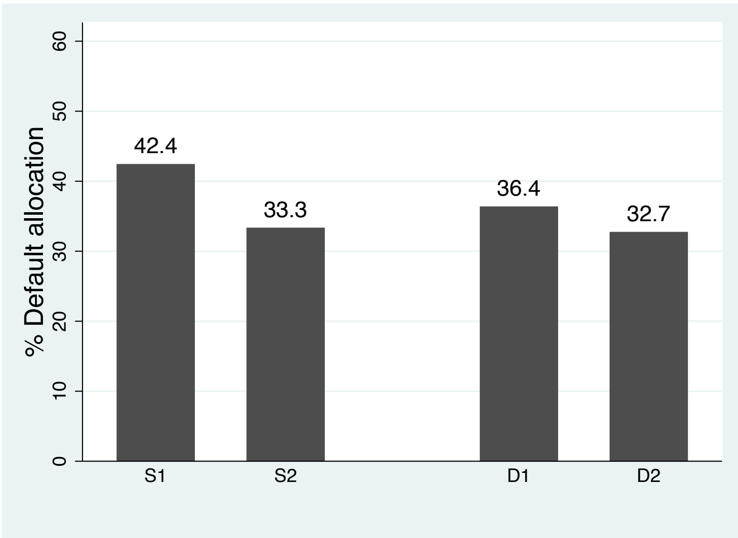


Figure 5: A-subjects' behavior by treatment

In line with Hypothesis 2.1, we find a negative correlation between the probability of choosing the default allocation and the expectation of acceptance (Spearman correlation coefficient equal to -0.34, p-value 0.00). Table 5 displays the t-tests of difference in means of the beliefs distribution for A-subjects across payoff treatments and methods of play. Both the t-test and the MW test indicate no significant difference in A-subjects' beliefs across payoff treatments (for fixed method of play). This means that we do not find evidence in favor of Hypothesis 2.3—i.e., that A-subjects' expectation of acceptance is higher in D1 than in D2.[28] Also, in line

[28]A-subjects have a correct estimate of the probability of acceptance in S1 (0.82 versus the actual share which is 0.73) and S2 (0.79 versus 0.79), while they overestimate the probability of

with B-subjects' behavior, A-subjects' expectations of acceptance are higher with the strategy than with the direct response method, significantly so in payoff treatment m_b^1. Yet, notwithstanding these first-order beliefs and the negative correlation between expectation of acceptance and probability of the default allocation, the frequency of default allocation does not differ across methods of play.

Table 5: A-subjects' expectation of acceptance

Mean α	m_b^1	m_b^2	Diff.	p-value (t-test)	p-value (MW)
Direct response	0.70	0.73	-0.03	0.59	0.46
Strategy	0.82	0.79	0.03	0.58	0.39
Diff.	-0.12	-0.06			
p-value (t-test)	0.03	0.23			
p-value (MW)	0.01	0.37			

Notes: means of A-subjects' subjective probability of acceptance, by payoff treatment and method of play; differences across treatments and p-values of a two-sided t-test of the differences in means and of a Mann-Whitney U test are displayed.

We estimate the effects of the treatments on A-subjects' behavior by considering a probit estimate of the probability of choosing the default allocation with Direct, Payoff Increase and their interaction as explanatory variables. As for B-subjects, we consider alternative specifications of the baseline probit model. In specification II we also consider the marginal effect of A-subjects' first-order beliefs about their co-player's move.[29] In specification III we control for individual observable characteristics (Age and Gender) and the subject's degree of morality—as elicited by the Aquino questionnaire—and the subject's degree of guilt aversion—as elicited by the GASP questionnaire.

The upper part of Table 6 reports the average marginal effects of the treatment dummies and the effect of their interaction, which is computed as the difference between the marginal effects of Payoff Increase with the two methods of play (reported in the lower part of the table) on the probability of choosing the default allocation, as implied by the probit estimate. Estimated marginal effects indicate that, contrary to Hypothesis 2.4, the payoff increase does not have any significant effect on the probability of the default allocation with either method of play. Also, differently from B-subjects and at odds with Hypothesis 2.2, A-subjects' behavior seems not to be affected by the method of play. Instead, in line with Hypothesis 2.1,

acceptance in D1 (0.70 versus 0.51) and D2 (0.73 versus 0.57).

[29]Unlike Table 4, we do not include the interaction Belief × Direct in specification II and III. Only for B-subjects the effect of such interaction was predicted by the theory. However, the reported estimates are robust to its inclusion.

probit estimates confirm that A-subjects' first-order beliefs have a significant negative effect on the probability of the default allocation. Morality and Guilt aversion seem not to affect A-subjects' behavior. The estimated marginal effect of Gender indicates that females are more likely to choose the default allocation than males.

Table 6: A-subjects' behavior

	I	II	III
Direct	-0.033	-0.100	-0.068
	(0.075)	(0.068)	(0.067)
Payoff Increase	-0.057	-0.052	-0.033
	(0.072)	(0.066)	(0.064)
Direct × Payoff Increase	0.055	0.101	0.101
	(0.149)	(0.136)	(0.131)
Belief		-0.008***	-0.007***
		(0.001)	(0.001)
Morality			0.051
			(0.060)
Guilt aversion			0.014
			(0.031)
Gender			0.189***
			(0.063)
Age			-0.016
			(0.022)
Payoff Increase if Direct = 1	-0.036	-0.014	0.006
	(0.091)	(0.080)	(0.080)
Payoff Increase if Direct = 0	-0.091	-0.115	-0.095
	(0.119)	(0.110)	(0.105)
Observations	176	176	176

Notes: the dependent variable is a dummy that indicates whether i chooses the default allocation; the table reports average marginal effects estimated from probit models; for the treatment interaction, we report the differential marginal effects of Payoff Increase when Direct changes from 0 to 1 (conditional marginal effects are reported in the lower part of the table); standard errors in parentheses; significance levels are: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Regressing A-subjects probability of choosing the default allocation on Payoff Increase unconditionally on the method of play delivers parameter estimates that are not statistically significant either controlling (average marginal effect: -0.057, p-value: 0.43), or not controlling for beliefs (average marginal effect: -0.052, p-value: 0.43). This result suggests that, differently from beliefs, inequity aversion, or other types of concern for the payoffs distribution, may not be an important driver of A-subjects' behavior.

Summarizing, our experimental results are as follows.

- **Result 1.1**: There is a significant negative correlation between the initial expectations of the default allocation and the probability of accepting the greedy offer with the direct response method, as predicted by Hypothesis 1.1, but we also find mild evidence of such correlation with the strategy method. Yet, the negative effect of initial expectations on the probability of acceptance is larger with the direct response method than with the strategy method.

- **Result 1.2**: The frequency of acceptance of the greedy offer is significantly lower with the direct response method than with the strategy method (Hypothesis 1.2 is verified).

- **Result 1.3**: B-subjects' average expectation of the default allocation in D2 is not lower than in D1 (Hypothesis 1.3 is verified).

- **Result 1.4**: The frequency of acceptance of the greedy offer is not significantly different both across S2 and S1 and across D1 and D2 (Hypothesis 1.4 is not verified).

- **Result 2.1**: In both methods of play, there is a significant negative correlation between the expectation of acceptance and the probability of choosing the default allocation (Hypothesis 2.1 is verified).

- **Result 2.2**: The frequency of the default allocation is not significantly different across methods of play (Hypothesis 2.2 is not verified).

- **Result 2.3**: A-subjects' average expectation of acceptance is not significantly different across payoff treatments, with both methods of play (Hypothesis 2.3 is not verified).

- **Result 2.4**: The frequency of the default allocation is not significantly different across payoff treatments, with both methods of play (Hypothesis 2.4 is not verified).

In a nutshell, the two experimental hypotheses about A and B-subjects' beliefs and behavior *within* treatments (Hypotheses 1.1 and 2.1) are both confirmed by the data. For what regards the comparison of behavior *across* treatments, the hypothesis about B-subjects' behavior across methods of play (Hypothesis 1.2) is verified (but not the one about A-subjects' behavior). Instead, the payoff manipulation does not seem to have the expected effect either on the frequency of acceptance or of default allocations. We discuss possible explanations of this result in Section 7.

7 Discussion

In line with the theory of frustration and anger we find a significant positive effect of initial beliefs on rejections with the direct response method (treatments D1 and D2). This contrasts with the observed effect of the payoff manipulation with the direct response method. Given that B-subjects' beliefs are on average similar in the two payoff treatments, B-subjects' initially expected payoff should be on average larger in the high payoff treatment D2 than in the baseline treatment D1. Yet, according to the data, rejections are not more frequent in D2 than in D1. This evidence suggests that there is another force that contrasts and neutralizes the predicted effect of the payoff increase on rejections.

A possible explanation relies on the fact that in the baseline treatment D1 the default allocation is an equal split, while in the high payoff treatment D2 it gives more to B. Thus, although in both cases the default allocation is the most upright choice, the greedy offer in D1 may be perceived by B-subjects as a more severe violation of a moral standard than in D2. When the perception of moral outrage is stronger, there may be less room for the frustration by goal obstruction to affect responses. In other words, some B-subjects might blame A-subjects more for being unfair than for decreasing their expected payoff, and thus may blame A-subjects more in D1 than in D2.

In Appendix 3 we report findings about A and B-subjects' behavior by gender. Since our experiment was not set out to test for gender differences,[30] any comparison across the females' and males' samples need to be considered with due caution. Yet, differences of behavior across genders may help interpret the absence of a payoff treatment effect. In a nutshell, while male B-subjects display a higher frequency of rejections in D2 than in D1 (although the difference is not significant), female B-subjects display a higher frequency of rejections in D1 than in D2. This evidence suggests that female B-subjects may be especially affected by anger as moral outrage.[31] Instead, males seem more sensitive to anger due to goal blockage. Moreover, evidence about A-subjects' behavior by gender suggests that female (male) subjects internalize the response of their female (male) counterpart. Thus, the opposite effect that the payoff increase has on B-subjects' behavior for males and females (with the direct response method) may be due to a gender-dependent appraisal of the greedy offer.

According to Social Role Theory (Eagly and Wood 2011), behavioral differences

[30] The number of females and males subjects differ—with a large majority of males—and we did not control for the gender distributions within sessions.

[31] Indeed, female subjects of our experiment scored higher than males in the Aquino questionnaire on moral identity.

between males and females derive from their different social roles. Males were originally the primary family providers, more concerned with power-based emotions and personal achievement. Females were originally the primary caretakers, more concerned with interpersonal relations and relational consequences.[32] Thus, evolutionary pressures may have led to a higher sensitivity to personal goal blockage for males and to violation of a social norm for females, with the related difference in emotional responses.

According to BDS theory, the responder cannot be frustrated with the strategy method, because he chooses a conditional response without observing whether the proposer has made the greedy offer. Such feature of the theory motivates the method of play manipulation in our experiment. Yet, we could in principle admit that frustration may also derive by the mere imagination of the unfavorable event of the greedy offer before choosing the conditional response. Consistently with this hypothesis, we find mild evidence of belief-dependence with the strategy method. We considered a generalized model assuming that frustration can, to some extent, be felt also with the strategy method (the analysis is available upon request). Notice that with the strategy method and imagined frustration, the responder's decision would depend on his first-order beliefs, whose distribution may differ from the one with the direct response method. In order to predict a higher rejection rate with the direct response method, it is sufficient to assume that B-subjects with the direct response method do not tend to attach a lower probability to the default allocation than with the strategy method. Table 2 shows mixed evidence about this assumption: it is verified in the high payoff treatments (m_{b}^2) and rejected at 10% in the baseline payoff treatments (m_{b}^1). However, this does not preclude a higher rejection rate with the direct response method than with the strategy method, given that the specified assumption is only sufficient.

We now consider another possible generalization of our model. Given our focus on responders' behavior, we simplified the analysis of proposers' assuming that they cannot be inequity averse. Yet, since proposers and responders are drawn from the same pool, it is reasonable to assume that also proposers may be inequity averse. A theoretical analysis of such generalization (available upon request) provides the following insights on proposers' behavior across treatments. Regarding the payoff-treatment comparison with the direct response method, there are two opposing forces. On one hand, we have the effect already analyzed in the paper: since A

[32] While two experimental studies on the Ultimatum Game provide mixed evidence about gender differences in responders' behavior (Eckel and Grossman 2001 and Solnick 2001), findings of several experiments on other social dilemmas support the view that women care more about equality (Eckel and Grossman 1998, Andreoni and Vesterlund 2001, Güth *et al.* 2007).

expects acceptance with higher probability in D1 than in D2, he is more willing to choose the default allocation in D2 than in D1. On the other hand, there is the effect we neglected, due to inequity aversion: A may be less inclined to choose the default allocation in D2 than in D1 because of the disadvantageous payoff distribution. The resulting relation between frequencies of the default allocation in the two treatments would depend on the distributions of proposers' personal features—i.e., the degree of inequity aversion and the treatment-specific beliefs. With the strategy method, instead, only inequity aversion is at work—as A's first-order beliefs do not vary with the payoff treatment—resulting in a lower frequency of the default allocation in S2 than in S1, which is at odds with our prediction and our findings. Instead, the result of Proposition 5 about the effect of the method of play on A's behavior would still hold, as the payoff distribution is constant across the direct response and the strategy method. The absence of a payoff treatment effect (with either method of play) in a probit estimation where we controlled for beliefs suggests that inequity aversion, unlike beliefs, is not an important driver of proposers' behavior.

In the rest of this section, we discuss how some of our main experimental and theoretical findings compare to those of previous studies.

As discussed in Section 2, in the experimental literature on the Ultimatum Game (e.g., Güth *et al.* 2001, Oxoby and McLeish 2004) there is mixed evidence on whether the method of play affects rejections—which is instead a clear result of our experiment. The meta-analysis by Oosterbeek *et al.* (2004) points at an opposite effect (higher rejection rate with the strategy method) with respect to ours.[33] Our model cannot explain this opposite result. Thus, we can only speculate on the reasons for such discrepancies.

An important feature of the meta-analysis is that it considers only Ultimatum Game experiments where the strategy method is implemented by asking responders their lowest acceptable offer, henceforth the *strategy threshold method*, and not by asking responders to make a choice contingent on every possible offer of the proposer, henceforth the *strategy vector method*. This can make a difference if we consider that boundedly rational responders may confuse covert with overt commitment—i.e., they may erroneously believe that their declared threshold is observed by the proposer before an offer is made, as it typically occurs in real life negotiations. Thus, using the strategy threshold method may trigger a framing effect: under the illusion of influencing the proposer's behavior, responders choose a threshold that is higher than the minimum offer they would in fact accept if they actually faced it. Instead, presenting responders with each hypothetical offer may force them to focus on what they would do if they actually received that specific offer.

[33]We thank the Associate Editor for bringing this paper to our attention.

The strategy threshold method might trigger another possible framing effect. While facing every hypothetical contingency may make the responder focus on his personal interest, being asked to decide a threshold can induce him to think more deeply at the moral standard.[34] Thus, while with the direct response method rejections are induced by frustration from not attaining a personal goal, with the strategy threshold method they may be induced by a commitment against morally unacceptable offers and may not truly reflect the responder's underlying propensity to accept or reject (given his personal traits). With this, a higher rejection rate with the strategy threshold method may obtain, even if the frustration motive in this condition is switched off.

An alternative motivation that can trigger aggressive behavior of the responder in the UMG is negative reciprocity (Rabin 1993, Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006): a rejection may be the reaction to the perceived unkindness of the proposer associated to the greedy offer.[35] While reciprocity *à la* Dufwenberg and Kirchsteiger (2004) has behavioral implications similar to our hypothesized payoff-treatment effect with the direct response method,[36] it does not provide a mechanism to expect a difference in behavior across methods of play. Furthermore, it yields different predictions about the correlation between responders' behavior and beliefs. The willingness to reject under reciprocity does not depend on the responder's first-order beliefs, but only on the perceived intentions of the proposer, and thus on the responder's second-order beliefs. Therefore, in contrast with BDS theory and our findings, reciprocity predicts zero correlation between responders' expectations of the default allocation and rejections also with the direct response method (assuming zero correlation between first- and second-order beliefs).

We discussed in Section 2 how cross-disciplinary findings support the relevance of initial expectations in negotiations and this is the main reason why we favor anger as a motivation of rejections. However, one can argue that both negative reciprocity and anger may simultaneously affect players' decisions in negotiations. We do not dismiss this hypothesis, which deserves further investigation.

[34]Blount and Bazerman (1996) find that responders are more demanding with the strategy threshold than with the strategy vector method and conjecture that the former induces subjects to pose greater attention to comparative rather than to absolute payoffs.

[35]Dufwenberg and Kirchsteiger (2014) discusses the role of reciprocity in the Ultimatum Game. Section 4 of Battigalli *et al.* (2019b) compares the implications of anger and reciprocity in the UMG and other games. Yet, these discussions rely on the equilibrium assumption.

[36]Reciprocity implies that increasing the responder's payoff in the default allocation mechanically increases the unkindness associated to the greedy offer, raising the incentive to reject. Unlike our model, this effects should kick in with both methods of play.

8 Conclusion

This paper presents a theory-driven experiment to test BDS theory of frustration and anger in the strategic context of the Ultimatum Minigame. From a theoretical perspective, a first contribution of this paper is the derivation of testable behavioral predictions without assuming either complete information, or equilibrium play. It is only assumed that incompletely informed players are subjectively rational (in a sense that encompasses psychological motivations), that proposers are confident in responders' subjective rationality, and that beliefs satisfy some mild across-treatment restrictions. The evidence that we find is mixed.

The experimental analysis delivers an important result in support of the theory: the higher the responder's expectation of the default allocation, the more likely it is that he rejects the greedy offer. This is in line with the theory of frustration and anger that assumes that the responder's frustration generated by the greedy offer, hence his tendency to reject, is increasing in his initially expected payoff.

Another important result, related to the first, regards the method of play. Rejections are more frequent with the direct response than with the strategy method, and with the direct response method the negative effect of initial expectations on the probability of accepting the greedy offer is stronger than with the strategy method. The differential effect of beliefs with the two methods of play provides additional support to the relevance of frustration and anger. Indeed, not only initial beliefs matter, but the fact that they matter more when responders can actually observe the greedy offer before making their choice confirms that frustration is a credible explanation for rejections. These results suggest that using the direct response method is more appropriate when observing previous moves is likely to induce emotional reactions.

Finally, we have extensively discussed the possible reasons why the payoff manipulation has no effect on aggregate behavior. Such discussion provides insights for future research. In particular, our analysis reveals that there might be alternative models of frustration and aggressive behavior, in which moral standards and gender play a significant role.

Appendix 1

Derivation of the incentive-compatible probability of acceptance Γ

Consider the UMG played with the direct response method. B expects to accept the greedy offer, i.e., $\gamma = 1$, if

$$\ell - h\theta F^{\mathrm{D},m_{\mathrm{b}}}(\beta, 1) - \delta(h - \ell) > 0,$$

that is,

$$\delta < \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta, 1, \theta) = \frac{\ell - h\theta\left(\beta m_{\mathrm{b}} + (1-\beta)\,\ell - \ell\right)}{h - \ell} = \frac{\ell - h\theta\beta(m_{\mathrm{b}} - \ell)}{h - \ell}.$$

Thus, the plan to accept is incentive compatible when anger and inequity aversion are low, i.e., for low values of β, θ, and δ. The yellow area in Figure 6 below represents the region of personal features (β, δ) satisfying (for a given θ) this incentive compatibility condition.

B instead expects to reject the greedy offer, i.e., $\gamma = 0$, if

$$\ell - h\theta F^{\mathrm{D},m_{\mathrm{b}}}(\beta, 0) - \delta(h - \ell) < 0,$$

that is,

$$\delta > \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta, 0, \theta) = \frac{\ell - h\theta \max\{0, \beta m_{\mathrm{b}} - \ell\}}{h - \ell}.$$

Thus, the plan to reject is incentive compatible when anger and inequity aversion are high, i.e., for high values of β, θ, and δ. The blue area in Figure 6 represents the region of personal features (β, δ) satisfying (for a given θ) this incentive compatibility condition.

Since $\max\{0, \beta m_{\mathrm{b}} - \ell\} \leq \beta\,(m_{\mathrm{b}} - \ell)$, we obtain the inequality between thresholds:

$$\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta, 1, \theta) = \frac{\ell - h\theta\beta(m_{\mathrm{b}} - \ell)}{h - \ell} \leq \frac{\ell - h\theta \max\{0, \beta m_{\mathrm{b}} - \ell\}}{h - \ell} = \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta, 0, \theta).$$

For intermediate values of β, θ, and δ, deterministic plans are self-defeating, i.e., a plan to accept induces rejection and a plan to reject induces acceptance *via* the effect of γ on the initially expected payoff and frustration. Thus, the only incentive-compatible plan is a probability of acceptance $\gamma \in (0, 1)$ that makes B indifferent. The indifference condition is

$$\ell - h\theta F^{\mathrm{D},m_{\mathrm{b}}}(\beta, \gamma) - \delta(h - \ell) = 0,$$

which holds for

$$\delta = \frac{\ell - h\theta F^{\mathrm{D},m_{\mathrm{b}}}(\beta, \gamma)}{h - \ell} \in \left[\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta, 1, \theta), \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta, 0, \theta)\right].$$

We pin down the incentive-compatible probability of acceptance γ as the solution of the following indifference condition:

$$\ell - h\theta\left(\beta m_{\mathrm{b}} + (1 - \beta)\gamma\ell - \ell\right) - \delta(h - \ell) = 0,$$

that is

$$\gamma = \frac{\ell - h\theta\left(\beta m_{\mathrm{b}} - \ell\right) - \delta\left(h - \ell\right)}{h\theta\left(1 - \beta\right)\ell} \text{ if } \theta > 0 \text{ and } 0 < \beta < 1.$$

The acceptance probability is *not* pinned down by the intrapersonal equilibrium condition if and only if $\delta = \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta, 1, \theta) = \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta, 0, \theta)$. The second equality holds if and only if either $\theta = 0$ (frustration does not matter), or $\beta = 0$ (no frustration), or $\beta = 1$ (the initially expected payoff is independent of γ).

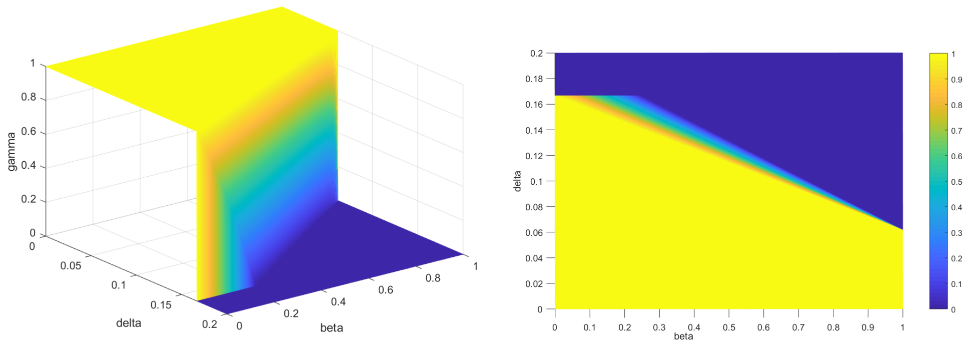


Figure 6: B's intrapersonal equilibrium correspondence as a function of β and δ, with parameters $\theta = 0.015, h = 14, m = 8, \ell = 2$.

Instead, with the strategy method, the probability of acceptance depends only on δ as shown in Proposition 1.

Neglecting non-generic cases that make γ indeterminate, we provide a unified representation of B's intrapersonal equilibrium function for the direct response and the strategy method as follows:

$$\Gamma^{\mathrm{P},m_{\mathrm{b}}}(\beta, \theta, \delta) = \begin{cases} 1 & \text{if } \delta < \hat{\delta}^{\mathrm{P},m_{\mathrm{b}}}(\beta, 1, \theta) \\ \frac{\ell - h\theta(\beta m_{\mathrm{b}} - \ell) - \delta(h-\ell)}{h\theta(1-\beta)\ell} & \text{if } \delta \in \left[\hat{\delta}^{\mathrm{P},m_{\mathrm{b}}}(\beta, 1, \theta), \hat{\delta}^{\mathrm{P},m_{\mathrm{b}}}(\beta, 0, \theta)\right] \\ 0 & \text{if } \delta > \hat{\delta}^{\mathrm{P},m_{\mathrm{b}}}(\beta, 0, \theta). \end{cases}$$

Proof of Proposition 2

Consider the incentive-compatible probability of acceptance with the direct response method $\Gamma^{\mathrm{D},m_{\mathrm{b}}}(\beta,\theta,\delta)$. Take β,β' such that $\beta' > \beta$. First, it is easy to check that $\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta',1,\theta) \leq \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta,1,\theta)$ and $\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta',0,\theta) \leq \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta,0,\theta)$. We have to consider the non-trivial case where $\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\cdot,1,\theta) < \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\cdot,0,\theta)$ for both β and β', and δ falls in the intermediate region: $\delta \in \left[\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta,1,\theta), \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta',0,\theta)\right]$.

We need to show that

$$\frac{\ell - h\theta\left(\beta' m_{\mathrm{b}} - \ell\right) - \delta(h-\ell)}{h\theta(1-\beta')\ell} \leq \frac{\ell - h\theta\left(\beta m_{\mathrm{b}} - \ell\right) - \delta(h-\ell)}{h\theta(1-\beta)\ell}$$

$$\Leftrightarrow (1-\beta)\left[\ell - h\theta\left(\beta' m_{\mathrm{b}} - \ell\right) - \delta(h-\ell)\right] \leq (1-\beta')\left[\ell - h\theta\left(\beta m_{\mathrm{b}} - \ell\right) - \delta(h-\ell)\right]$$

$$\Leftrightarrow (\beta' - \beta)\left[\ell - h\theta\left(m_{\mathrm{b}} - \ell\right) - \delta(h-\ell)\right] \leq 0.$$

By assumption $(\beta' - \beta) > 0$. The term in brackets is negative because:

$$\ell - h\theta\left(m_{\mathrm{b}} - \ell\right) - \delta(h-\ell) \leq \ell - h\theta\left(\beta m_{\mathrm{b}} + (1-\beta)\ell - \ell\right) - \delta(h-\ell) \leq 0,$$

where the first inequality follows from $m_{\mathrm{b}} > \ell$ and the second from $\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta,1,\theta) \leq \delta \leq \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta',0,\theta)$; indeed,

$$\delta \geq \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta,1,\theta) \Leftrightarrow \delta \geq \frac{\ell - h\theta\left(\beta m_{\mathrm{b}} + (1-\beta)\ell - \ell\right)}{h-\ell}$$

$$\delta(h-\ell) \geq \ell - h\theta\left(\beta m_{\mathrm{b}} + (1-\beta)\ell - \ell\right)$$

$$\ell - h\theta\left(\beta m_{\mathrm{b}} + (1-\beta)\ell - \ell\right) - \delta(h-\ell) \leq 0,$$

where the latter inequality is implied by the indifference condition.

Instead, the incentive-compatible probability of acceptance with the strategy method $\Gamma^{\mathrm{S},m_{\mathrm{b}}}$ is independent of β, since, as shown in Proposition1, the acceptance condition with the strategy method depends only on δ. ■

Appendix 2

Behavioral predictions about responders

For each treatment, we posit a distribution of the relevant personal features of B-subject (β, θ, δ). Specifically, let $\mu^{\mathrm{Dk}} \in \Delta\left([0,1] \times \mathbb{R}_{+}^{2}\right)$ denote the distribution of B's personal features in treatment Dk, with $\mathrm{k} \in \{1,2\}$ indicating the payoff condition $m_{\mathrm{b}}^{\mathrm{k}}$. We assume μ^{Dk} to be atomless. Recall that while the marginal distribution of personal traits (θ, δ) is exogenous—i.e., independent of the treatment—beliefs are endogenous, as they are affected by strategic thinking. Given the random draw of individuals (experimental subjects) to play the UMG in a specific role and treatment, the personal features of the individual playing in role B are vectors of random variables, which we denote with **boldface** letters. When considering inequalities among random variables, we interpret them as holding for almost every realization of the two random variables.[37] Thus, $\left(\boldsymbol{\beta}^{\mathrm{Pk}}, \boldsymbol{\theta}, \boldsymbol{\delta}\right)$ is the random vector of B's personal features, where $\boldsymbol{\beta}^{\mathrm{Pk}}$ represents B's beliefs in treatment Pk. Moreover, let $\boldsymbol{\gamma}^{\mathrm{Pk}}$ denote B's planned probability of acceptance of the greedy offer in treatment Pk.

The analysis B-subjects is based on (i) subjective rationality, that is, B's behavior strategy is incentive-compatible and B actually chooses according to such behavior strategy, and (ii) an auxiliary assumption about B-subjects' beliefs in D1 and D2.

Assumption 1. *(i) Player B is subjectively rational, i.e., he plans rationally given his beliefs about the other, and implements his plan; (ii) B's belief distributions in D1 and D2 satisfy* $\boldsymbol{\beta}^{\mathrm{D2}} \geq \boldsymbol{\beta}^{\mathrm{D1}}$.

Recall that $\Gamma^{\mathrm{Pk}}(\beta, \theta, \delta)$ denotes the incentive-compatible probability of accepting the greedy offer given personal features (β, θ, δ) in treatment Pk. Due to the intrapersonal equilibrium assumption, the objective probability that B accepts the greedy offer coincides with the subjective probability that B attaches to this event. With this, the frequency of acceptance of the greedy offer is

$$\mathbb{P}\left[\mathbf{a}_{\mathrm{b}}^{\mathrm{Pk}} = y\right] = \mathbb{E}[\boldsymbol{\gamma}^{\mathrm{Pk}}] = \int \boldsymbol{\Gamma}^{\mathrm{Pk}} \mathrm{d}\mu^{\mathrm{Pk}}.$$

Proposition 3. *The frequency of acceptance of the greedy offer in treatment D1 is higher than in treatment D2, i.e.,*

$$\mathbb{P}[\mathbf{a}_{\mathrm{b}}^{\mathrm{D1}} = y] \geq \mathbb{P}[\mathbf{a}_{\mathrm{b}}^{\mathrm{D2}} = y].$$

[37] That is, the set of the realizations for which the inequality does not hold has zero Lebesgue measure.

Proof of Proposition 3. It is easily verified that Γ is decreasing in m_{b}:

$$\frac{\partial}{\partial m_{\mathrm{b}}}\Gamma^{\mathrm{Dk}}(\beta,\theta,\delta)=\begin{cases}-\frac{h\theta\beta}{h\theta(1-\beta)\ell} & \text{if } \delta\in\left[\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta,1,\theta),\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta,0,\theta)\right], \\ & \quad\text{with } \hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta,1,\theta)\neq\hat{\delta}^{\mathrm{D},m_{\mathrm{b}}}(\beta,0,\theta) \\ 0 & \text{otherwise.}\end{cases}$$

With this, since $m_{\mathrm{b}}^{1}<m_{\mathrm{b}}^{2}$, for all (β,θ,δ) it holds that

$$\Gamma^{\mathrm{D1}}(\beta,\theta,\delta)\geq\Gamma^{\mathrm{D2}}(\beta,\theta,\delta).$$

Hence

$$\mathbb{E}\left[\boldsymbol{\gamma}^{\mathrm{D1}}\right]=\int\boldsymbol{\Gamma}^{\mathrm{D1}}\mathrm{d}\mu^{\mathrm{D1}}\geq\int\boldsymbol{\Gamma}^{\mathrm{D2}}\mathrm{d}\mu^{\mathrm{D1}}.$$

Consider any version of the conditional probability measures $\mu^{\mathrm{Dk}}\left(\cdot|\boldsymbol{\theta},\boldsymbol{\delta}\right)$ (k=1,2). Since, due to Proposition 2, Γ^{D2} is decreasing in β, and we assumed $\boldsymbol{\beta}^{\mathrm{D2}}\geq\boldsymbol{\beta}^{\mathrm{D1}}$, it follows that

$$\int\Gamma^{\mathrm{D2}}\left(\beta,\boldsymbol{\theta},\boldsymbol{\delta}\right)\mu^{\mathrm{D1}}\left(\mathrm{d}\beta|\boldsymbol{\theta},\boldsymbol{\delta}\right)\geq\int\Gamma^{\mathrm{D2}}\left(\beta,\boldsymbol{\theta},\boldsymbol{\delta}\right)\mu^{\mathrm{D2}}\left(\mathrm{d}\beta|\boldsymbol{\theta},\boldsymbol{\delta}\right).$$

Therefore, taking into account that the marginal distribution of $(\boldsymbol{\theta},\boldsymbol{\delta})$ is the same in the two treatments because they are personal traits (hence exogenous), we obtain

$$\begin{aligned}\int\boldsymbol{\Gamma}^{\mathrm{D2}}\mathrm{d}\mu^{\mathrm{D\,1}} &= \int\left(\int\Gamma^{\mathrm{D2}}\left(\beta,\theta,\delta\right)\mu^{\mathrm{D1}}\left(\mathrm{d}\beta|\theta,\delta\right)\right)\mathrm{marg}_{\boldsymbol{\theta},\boldsymbol{\delta}}\mu^{\mathrm{D1}}\left(\mathrm{d}\theta,\mathrm{d}\delta\right)\\ &= \int\left(\int\Gamma^{\mathrm{D2}}\left(\beta,\theta,\delta\right)\mu^{\mathrm{D1}}\left(\mathrm{d}\beta|\theta,\delta\right)\right)\mathrm{marg}_{\boldsymbol{\theta},\boldsymbol{\delta}}\mu^{\mathrm{D2}}\left(\mathrm{d}\theta,\mathrm{d}\delta\right)\\ &\geq \int\left(\int\Gamma^{\mathrm{D2}}\left(\beta,\theta,\delta\right)\mu^{\mathrm{D2}}\left(\mathrm{d}\beta|\theta,\delta\right)\right)\mathrm{marg}_{\boldsymbol{\theta},\boldsymbol{\delta}}\mu^{\mathrm{D2}}\left(\mathrm{d}\theta,\mathrm{d}\delta\right)\\ &= \int\boldsymbol{\Gamma}^{\mathrm{D2}}\mathrm{d}\mu^{\mathrm{D2}}.\end{aligned}$$

Hence

$$\mathbb{E}\left[\boldsymbol{\gamma}^{\mathrm{D1}}\right]\geq\int\boldsymbol{\Gamma}^{\mathrm{D2}}\mathrm{d}\mu^{\mathrm{D1}}\geq\int\boldsymbol{\Gamma}^{\mathrm{D2}}\mathrm{d}\mu^{\mathrm{D2}}=\mathbb{E}\left[\boldsymbol{\gamma}^{\mathrm{D2}}\right],$$

as desired. ■

Proposition 3 is robust to the substitution of the intrapersonal equilibrium assumption with an alternative assumption on beliefs.[38]

Remark 1. *If $m_b^2 - m_b^1 > \ell$, i.e., the payoff increase is larger than player B's payoff from accepting the greedy offer, and $\boldsymbol{\beta}^{D2} \geq \boldsymbol{\beta}^{D1} \geq \frac{1}{2}$, then the frequency of acceptance of the greedy offer in treatment D1 is higher than in treatment D2.*

Proof of Remark 1. Denote with $\mathbf{M}_b^{Dk}$ player B's initially expected monetary payoff in treatment Dk, i.e.,

$$\mathbf{M}_b^{Dk}\left(\boldsymbol{\beta}^{Dk}, \boldsymbol{\gamma}^{Dk}\right) = \mathbb{E}_{\boldsymbol{\beta}^{Dk},\boldsymbol{\gamma}^{Dk}}\left[\mathbf{m}_b^k\right] = \beta^{Dk} m_b^k + (1-\beta^{Dk})\gamma^{Dk}\ell.$$

We first show that $\mathbf{M}_b^{D2}\left(\boldsymbol{\beta}^{D2}, \boldsymbol{\gamma}^{D2}\right) \geq \mathbf{M}_b^{D1}\left(\boldsymbol{\beta}^{D1}, \boldsymbol{\gamma}^{D1}\right)$ almost everywhere. Since $\mathrm{M}_b^{Dk}(\beta^{Dk}, \gamma^{Dk})$ is strictly increasing in γ^{Dk}, the following relation

$$\mathrm{M}_b^{D2}(\beta^{D2}, \gamma^{D2}) = \beta^{D2} m_b^2 + (1-\beta^{D2})\gamma^{D2}\ell \geq \beta^{D1} m_b^1 + (1-\beta^{D1})\gamma^{D1}\ell = \mathrm{M}_b^{D1}(\beta^{D1}, \gamma^{D1})$$

holds for all γ^{D1} and γ^{D2} if and only if it holds with $\gamma^{D1} = 1$ and $\gamma^{D2} = 0$, i.e., if and only if

$$\beta^{D2} m_b^2 \geq \beta^{D1} m_b^1 + \left(1 - \beta^{D1}\right)\ell.$$

Since for all β^{D1} and β^{D2} it holds that $\beta^{D2} \geq \beta^{D1} \geq \frac{1}{2} \geq (1-\beta^{D1})$ and $(m_b^2 - m_b^1) > \ell$, we obtain

$$\begin{aligned}\beta^{D2} m_b^2 - \beta^{D1} m_b^1 \geq \beta^{D1} m_b^2 - \beta^{D1} m_b^1 &= \beta^{D1}\left(m_b^2 - m_b^1\right) \\ &\geq \frac{1}{2}\left(m_b^2 - m_b^1\right) > \frac{1}{2}\ell \geq \left(1 - \beta^{D1}\right)\ell,\end{aligned}$$

as desired.

With this, we now show that the frequency of acceptance of the greedy offer in treatment D1 is higher than in treatment D2. Player B with personal features $(\beta, \gamma, \theta, \delta)$ accepts the greedy offer in treatment Dk if and only if

$$\mathrm{M}_b^{Dk}(\beta, \gamma) \leq \hat{m}(\theta, \delta) := \frac{\ell + \theta\ell h - \delta(h - \ell)}{\theta h}$$

Thus, we can compute the frequency of accepting the greedy offer in treatment Dk as:

$$\mathbb{P}\left[\mathbf{a}_\mathbf{b}^{Dk} = y\right] = \int \mathbb{1}_{\mathrm{M}_b^{Dk}(\beta,\gamma) \leq \hat{m}(\theta,\delta)} \mu^{Dk}(\mathrm{d}\beta, \mathrm{d}\gamma, \mathrm{d}\theta, \mathrm{d}\delta)$$

[38] Without the intrapersonal equilibrium assumption, the larger payoff from the default allocation does not necessarily imply that B-subjects in D2 are more frustrated than B-subjects in D1. To guarantee that, we make an alternative assumption on B's beliefs. This implies that B's expected payoff in D1 is higher than in D2 for every realization of $\boldsymbol{\gamma}^{D1}$ and $\boldsymbol{\gamma}^{D2}$ even in the worst case scenario, that is where the expectation of accepting the greedy offer in D1 (D2) is highest (lowest) and thus the expected payoff in D1 (D2) is maximal (minimal) ($\gamma^{D1} = 1$ and $\gamma^{D2} = 0$).

where $\mu^{\mathrm{Dk}} \in \Delta\left([0,1]^2 \times \mathbb{R}_+^2\right)$ is the treatment-dependent measure on all the personal features of B, including plan $\boldsymbol{\gamma}$.

To obtain $\mathbb{P}\left[\mathbf{a}_{\mathbf{b}}^{\mathrm{D1}} = y\right] \geq \mathbb{P}\left[\mathbf{a}_{\mathbf{b}}^{\mathrm{D2}} = y\right]$, we first have to show that $\mathbb{P}\left[\mathbf{a}_{\mathbf{b}}^{\mathrm{D1}} = y|\boldsymbol{\theta},\boldsymbol{\delta}\right] \geq \mathbb{P}\left[\mathbf{a}_{\mathbf{b}}^{\mathrm{D2}} = y|\boldsymbol{\theta},\boldsymbol{\delta}\right]$, i.e., that

$$\int \mathbb{1}_{\mathrm{M}_{\mathrm{b}}^{\mathrm{D1}}(\beta,\gamma)\leq \hat{m}(\boldsymbol{\theta},\boldsymbol{\delta})}\mu^{\mathrm{D1}}(\mathrm{d}\beta,\mathrm{d}\gamma|\boldsymbol{\theta},\boldsymbol{\delta}) \geq \int \mathbb{1}_{\mathrm{M}_{\mathrm{b}}^{\mathrm{D2}}(\beta,\gamma)\leq \hat{m}(\boldsymbol{\theta},\boldsymbol{\delta})}\mu^{\mathrm{D2}}(\mathrm{d}\beta,\mathrm{d}\gamma|\boldsymbol{\theta},\boldsymbol{\delta}).$$

Indeed, since $\mathbf{M}_{\mathrm{b}}^{\mathrm{D2}} \geq \mathbf{M}_{\mathrm{b}}^{\mathrm{D1}}$ almost everywhere, $\boldsymbol{\beta}^{\mathrm{D2}} \geq \boldsymbol{\beta}^{\mathrm{D1}}$, and $\mathbf{M}_{\mathrm{b}}^{\mathrm{Dk}}$ is decreasing in β, it follows that

$$\begin{aligned}\int \mathbb{1}_{\mathrm{M}_{\mathrm{b}}^{\mathrm{D1}}(\beta,\gamma)\leq \hat{m}(\boldsymbol{\theta},\boldsymbol{\delta})}\mu^{\mathrm{D1}}(\mathrm{d}\beta,\mathrm{d}\gamma|\boldsymbol{\theta},\boldsymbol{\delta}) &\geq \int \mathbb{1}_{\mathrm{M}_{\mathrm{b}}^{\mathrm{D2}}(\beta,\gamma)\leq \hat{m}(\boldsymbol{\theta},\boldsymbol{\delta})}\mu^{\mathrm{D1}}(\mathrm{d}\beta,\mathrm{d}\gamma|\boldsymbol{\theta},\boldsymbol{\delta}) \\ &\geq \int \mathbb{1}_{\mathrm{M}_{\mathrm{b}}^{\mathrm{D2}}(\beta,\gamma)\leq \hat{m}(\boldsymbol{\theta},\boldsymbol{\delta})}\mu^{\mathrm{D2}}(\mathrm{d}\beta,\mathrm{d}\gamma|\boldsymbol{\theta},\boldsymbol{\delta}).\end{aligned}$$

Therefore, taking into account that the marginal distribution of $(\boldsymbol{\theta},\boldsymbol{\delta})$ is the same in the two treatments because they are personal traits (hence exogenous), we obtain

$$\begin{aligned}\mathbb{P}\left[\mathbf{a}_{\mathbf{b}}^{\mathrm{D1}} = y\right] &= \int \left(\int \mathbb{1}_{\mathrm{M}_{\mathrm{b}}^{\mathrm{D1}}(\beta,\gamma)\leq \hat{m}(\theta,\delta)}\mu^{\mathrm{D1}}(\mathrm{d}\beta,\mathrm{d}\gamma|\theta,\delta)\right)\mathrm{marg}_{\boldsymbol{\theta},\boldsymbol{\delta}}\mu^{\mathrm{D1}}(\mathrm{d}\theta,\mathrm{d}\delta) \\ &\geq \int \left(\int \mathbb{1}_{\mathrm{M}_{\mathrm{b}}^{\mathrm{D2}}(\beta,\gamma)\leq \hat{m}(\theta,\delta)}\mu^{\mathrm{D2}}(\mathrm{d}\beta,\mathrm{d}\gamma|\theta,\delta)\right)\mathrm{marg}_{\boldsymbol{\theta},\boldsymbol{\delta}}\mu^{\mathrm{D2}}(\mathrm{d}\theta,\mathrm{d}\delta) \\ &= \mathbb{P}\left[\mathbf{a}_{\mathbf{b}}^{\mathrm{D2}} = y\right],\end{aligned}$$

as desired. ■

We now make a comparative prediction about B's behavior across methods of play.

Proposition 4. *For each payoff treatment, the frequency of acceptance of the greedy offer with the strategy method is higher than the conditional frequency of acceptance of the greedy offer with the direct response method, i.e., for $k = 1, 2$,*

$$\mathbb{P}[\mathbf{a}_{\mathrm{b}}^{\mathrm{Sk}} = y] \geq \mathbb{P}[\mathbf{a}_{\mathrm{b}}^{\mathrm{Dk}} = y|\mathbf{a}_{\mathrm{a}}^{\mathrm{Dk}} = g].$$

Proof of Proposition 4. It is easily verified that

$$\mathbb{E}\left[\boldsymbol{\gamma}^{\mathrm{Dk}}\right] = \int \boldsymbol{\Gamma}^{\mathrm{Dk}}\mathrm{d}\mu^{\mathrm{Dk}} \leq \int \boldsymbol{\Gamma}^{\mathrm{Sk}}\mathrm{d}\mu^{\mathrm{Dk}},$$

because $\Gamma^{\mathrm{Dk}}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\delta}) \leq \Gamma^{\mathrm{Sk}}(\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\delta})$ almost everywhere. To see this, notice that

$\Gamma^{\mathrm{Sk}} = 0$ implies $\Gamma^{\mathrm{Dk}} = 0$ because

$$\hat{\delta}^{\mathrm{Sk}}(\beta, 0, \theta) = \frac{\ell}{h - \ell} \geq \frac{\ell - h\theta \max\{0, \beta m_{\mathrm{b}} - \ell\}}{h - \ell} = \hat{\delta}^{\mathrm{Dk}}(\beta, 0, \theta)$$

for all β and θ. Note that B's probability of acceptance in treatment Sk, Γ^{Sk}, depends only on the exogenous personal trait δ, hence it is treatment-independent:

$$\int \mathbf{\Gamma}^{\mathrm{Sk}} \mathrm{d}\mu^{\mathrm{Dk}} = \int \mathbf{\Gamma}^{\mathrm{Sk}} \mathrm{d}\mu^{\mathrm{Sk}}.$$

Hence,

$$\mathbb{E}\left[\boldsymbol{\gamma}^{\mathrm{Dk}}\right] \leq \int \mathbf{\Gamma}^{\mathrm{Sk}} \mathrm{d}\mu^{\mathrm{Sk}} = \mathbb{E}\left[\boldsymbol{\gamma}^{\mathrm{Sk}}\right],$$

as desired. ■

Behavioral predictions about proposers

Given the simplifying assumption that A maximizes his subjective expected payoff, the relevant personal features of the individual playing in role A coincide with his personal beliefs: his initial first-order beliefs about B accepting the greedy offer and his second-order beliefs about B's initial expectation of a default allocation. Denote by ϕ the marginal second-order belief of A about B's personal features, i.e., a joint belief on B's first-order belief β and personal traits (θ, δ) given by a subjective probability measure on $[0, 1] \times \mathbb{R}^2_+$.[39] We will show how A's first-order belief α is derived from ϕ.

Like we did for B-subjects, we posit a distribution of the relevant personal features of A-subjects (α, ϕ) for each treatment and we denote the random vector of A's personal features is $\left(\boldsymbol{\alpha}^{\mathrm{Pk}}, \boldsymbol{\phi}^{\mathrm{Pk}}\right)$.

The analysis A-subjects is based on (i) subjective rationality (subjective expected payoff maximization) given certainty of B's subjective rationality, and (ii) an auxiliary assumption about the second-order beliefs of A-subjects in treatments D1 and D2.

Assumption 2. *(i) Player A is subjectively rational and is certain that B is subjectively rational and (ii) A's belief distributions are such that, for any conditional*

[39]Measure $\phi \in \Delta\left([0, 1] \times \mathbb{R}^2_+\right)$ is a *marginal* second-order belief because A is assumed to hold a joint belief about B's *action* and personal features (β, θ, δ), which is the whole second-order belief of A.

probability measure $\phi^{\mathrm{Dk}}(\cdot|\boldsymbol{\theta},\boldsymbol{\delta})$ *(*$\mathrm{k}=1,2$*), the following holds almost everywhere*

$$\mathbb{P}_{\phi^{\mathrm{D2}}}\left[\boldsymbol{\beta}^{\mathbf{D2}} \leq \beta|\boldsymbol{\theta},\boldsymbol{\delta}\right] \leq \mathbb{P}_{\phi^{\mathrm{D1}}}\left[\boldsymbol{\beta}^{\mathbf{D1}} \leq \beta|\boldsymbol{\theta},\boldsymbol{\delta}\right] \text{ for all } \beta \in [0,1], \tag{5}$$

Given that A is certain of B's rationality, we obtain his first-order belief on conditional acceptance of the greedy offer as the expected acceptance rate calculated according to his personal second-order beliefs:

$$\alpha^{\mathrm{Pk}} = \mathbb{E}_{\phi^{\mathrm{Pk}}}\left[\boldsymbol{\Gamma}^{\mathrm{Pk}}\right] = \int \Gamma^{\mathrm{Pk}}(\beta,\theta,\delta)\,\phi^{\mathrm{Pk}}(\mathrm{d}\beta,\mathrm{d}\theta,\mathrm{d}\delta).$$

Lemma 1. *A's expectation that B accepts the greedy offer is lower in treatment D2 than in treatment D1, that is,*

$$\boldsymbol{\alpha}^{\mathrm{D1}} \geq \boldsymbol{\alpha}^{\mathrm{D2}}.$$

Proof of Lemma 1. We have to show that

$$\int \boldsymbol{\Gamma}^{\mathrm{D1}}\mathrm{d}\phi^{\mathrm{D1}} \geq \int \boldsymbol{\Gamma}^{\mathrm{D2}}\mathrm{d}\phi^{\mathrm{D2}}.$$

Since $m_{\mathrm{b}}^2 > m_{\mathrm{b}}^1$ implies that $\Gamma^{\mathrm{D1}}(\beta,\theta,\delta) \geq \Gamma^{\mathrm{D2}}(\beta,\theta,\delta)$ for all (β,θ,δ), it follows that

$$\int \boldsymbol{\Gamma}^{\mathrm{D1}}\mathrm{d}\phi^{\mathrm{D1}} \geq \int \boldsymbol{\Gamma}^{\mathrm{D2}}\mathrm{d}\phi^{\mathrm{D1}}.$$

Then,

$$\begin{aligned}
\int \boldsymbol{\Gamma}^{\mathrm{D2}}\mathrm{d}\phi^{\mathrm{D1}} &= \int\left(\int \Gamma^{\mathrm{D2}}(\beta,\theta,\delta)\,\phi^{\mathrm{D1}}(\mathrm{d}\beta|\theta,\delta)\right)\mathrm{marg}_{\boldsymbol{\theta},\boldsymbol{\delta}}\phi^{\mathrm{D1}}(\mathrm{d}\theta,\mathrm{d}\delta)\\
&= \int\left(\int \Gamma^{\mathrm{D2}}(\beta,\theta,\delta)\,\phi^{\mathrm{D1}}(\mathrm{d}\beta|\theta,\delta)\right)\mathrm{marg}_{\boldsymbol{\theta},\boldsymbol{\delta}}\phi^{\mathrm{D2}}(\mathrm{d}\theta,\mathrm{d}\delta)\\
&\geq \int\left(\int \Gamma^{\mathrm{D2}}(\beta,\theta,\delta)\,\phi^{\mathrm{D2}}(\mathrm{d}\beta|\theta,\delta)\right)\mathrm{marg}_{\boldsymbol{\theta},\boldsymbol{\delta}}\phi^{\mathrm{D2}}(\mathrm{d}\theta,\mathrm{d}\delta)\\
&= \int \boldsymbol{\Gamma}^{\mathrm{D2}}\mathrm{d}\phi^{\mathrm{D2}},
\end{aligned}$$

where the second equality holds because personal traits $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ are treatment-independent, and the inequality follows from Assumption 2 and the fact that Γ is decreasing in β by Proposition 2. ■

Proposition 5. *The frequency of the default allocation in treatment D2 is higher than in treatment D1, i.e.,*

$$\mathbb{P}[\mathbf{a}_{\mathrm{a}}^{\mathrm{D2}} = d] \geq \mathbb{P}[\mathbf{a}_{\mathrm{a}}^{\mathrm{D1}} = d].$$

Proof of Proposition 5. By Lemma 1, $\boldsymbol{\alpha}^{\mathrm{D1}} \geq \boldsymbol{\alpha}^{\ \mathrm{D2}}$. This implies that

$$\mathbb{P}\left[\boldsymbol{\alpha}^{\mathrm{D1}} \leq \hat{\alpha}\right] \leq \mathbb{P}\left[\boldsymbol{\alpha}^{\mathrm{D2}} \leq \hat{\alpha}\right],$$

i.e., it is less likely that a randomly drawn A-subject from treatment D1 has first-order belief $\alpha \leq \hat{\alpha}$ than a randomly drawn A-subject from treatment D2. Thus, due to incentive compatibility, it is more likely to draw an A-subject that chooses the default allocation in treatment D2 than in treatment D1. ■

We now make a comparative prediction about A's behavior across methods of play.

Proposition 6. *For each payoff treatment, the frequency of the default allocation with the direct response method is higher than with the strategy method, i.e., for* $k = 1, 2$,

$$\mathbb{P}\left[\mathbf{a}_{\mathrm{a}}^{\mathrm{Dk}} = d\right] \geq \mathbb{P}\left[\mathbf{a}_{\mathrm{a}}^{\mathrm{Sk}} = d\right].$$

Proof of Proposition 6. First, we show that, given the payoff treatment k, A's expectation that B accepts the greedy offer is lower in treatment Dk than in treatment Sk. Indeed, for $\mathrm{k} = 1, 2$ it holds that

$$\alpha^{\mathrm{Dk}} = \int \boldsymbol{\Gamma}^{\mathrm{Dk}} \mathrm{d}\phi^{\mathrm{Dk}} \leq \int \boldsymbol{\Gamma}^{\mathrm{Sk}} \mathrm{d}\phi^{\mathrm{Dk}} = \int \boldsymbol{\Gamma}^{\mathrm{Sk}} \mathrm{d}\phi^{\ \mathrm{Sk}} = \alpha^{\mathrm{Sk}}$$

where the inequality holds because $\Gamma^{\mathrm{Dk}}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta}) \leq \Gamma^{\mathrm{Sk}}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\delta})$ almost everywhere and the second equality holds because Γ^{Sk} depends only on the distribution of $\boldsymbol{\delta}$, which is exogenous, hence treatment-independent. Therefore, for each k it holds that $\boldsymbol{\alpha}^{\mathrm{Sk}} \geq \boldsymbol{\alpha}^{\mathrm{Dk}}$ almost everywhere, which implies that

$$\mathbb{P}\left[\boldsymbol{\alpha}^{\mathrm{Sk}} \leq \hat{\alpha}\right] \leq \mathbb{P}\left[\boldsymbol{\alpha}^{\mathrm{Dk}} \leq \hat{\alpha}\right],$$

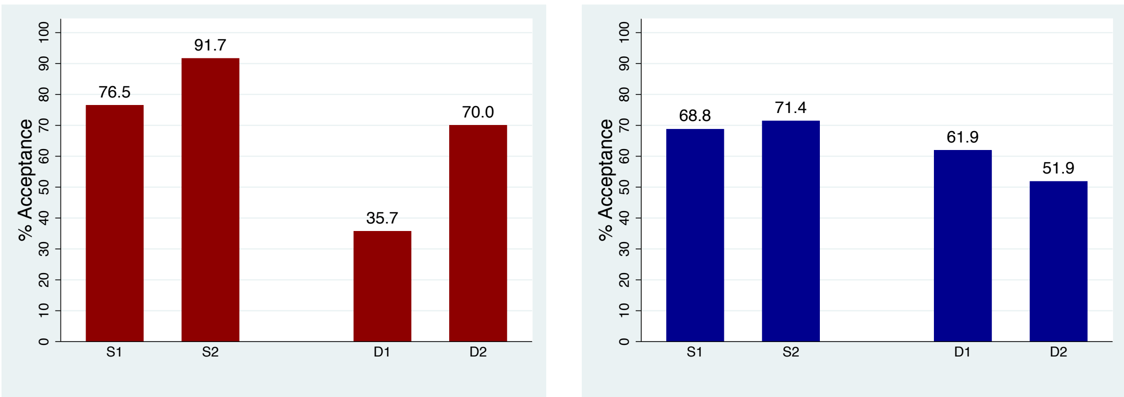
i.e., it is less likely that a randomly drawn A-subject from treatment Sk has first-order belief $\alpha \leq \hat{\alpha}$ than a randomly drawn A-subject from treatment Dk. Thus, due to incentive compatibility, it is more likely to draw an A-subject that chooses the default allocation in treatment Dk than in treatment Sk. ■

Appendix 3

Results by gender

This appendix reports gender specific results. Among B-subjects we had 112 males and 64 females. With the strategy method we had 37 males and 29 females, while with the direct response method, out of 72 B-subjects who received the greedy offer, 48 were males and 24 females. Among A-subjects we had 101 males and 75 females.

Figure 7 illustrates acceptance rates by method of play and payoff treatment across genders. As for females, the average frequency of acceptance is lower with the direct response method than with the strategy method (MW: p-value 0.01, 53 observations); there is mild evidence that the frequency of acceptance is *higher* in D2 than in D1 (MW: p-value 0.11; 24 observations) but there is no difference across S2 and S1 (MW: p-value 0.19, 29 observations). Considering males, the average frequency of acceptance is slightly lower with the direct response method than with the strategy method (MW: p-value 0.19, 85 observations) but it is similar across payoff treatments, for fixed method of play (MW with the direct response method: p-value 0.49, 48 observations; MW with the strategy method: p-value 0.86, 37 observations).



Females Males

Figure 7: B-subjects' behavior by gender

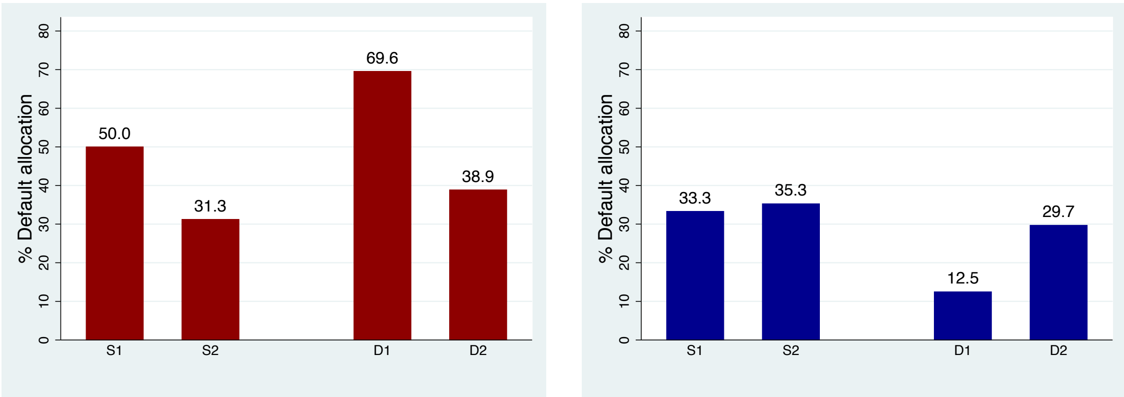
Table 7 reports separately for each gender the average marginal effects of the treatment dummies on the probability of accepting the greedy offer resulting from a probit estimate of the model described in equation (4). In particular, we consider specifications I and III of Table 4 in the main text. Examining the average marginal effect of Direct in both specifications, it results that the effect of the direct response method that we find for the whole sample is similar across genders: for both females and males, playing the direct response method rather than the strategy method decreases the probability of acceptance. Interestingly, the payoff treatment effect with the direct response method differs across genders. Specifically, Payoff Increase has a positive effect on the probability of acceptance for females and a negative effect on that of males. Yet, such conditional effect is significant only for females. Instead, with the strategy method, Payoff Increase has a small, and not significant, marginal effect on the probability of acceptance both for females and males.

Table 7: B-subjects' behavior by gender

	Females		Males	
	I	II	I	II
Direct	-0.328***	-0.354***	-0.140	-0.232**
	(0.118)	(0.113)	(0.104)	(0.093)
Payoff Increase	0.238**	0.195*	-0.045	-0.017
	(0.113)	(0.114)	(0.104)	(0.095)
Direct × Payoff Increase	0.191	0.260	-0.127	0.085
	(0.233)	(0.226)	(0.209)	(0.183)
Belief		-0.006**		-0.007***
		(0.003)		(0.002)
Belief × Direct		-0.002		-0.004
		(0.005)		(0.003)
Trait Anger		-0.003		-0.008
		(0.005)		(0.006)
Age		-0.008		-0.070**
		(0.029)		(0.028)
Payoff Increase if Direct = 1	0.343*	0.335*	-0.101	-0.054
	(0.193)	(0.185)	(0.143)	(0.125)
Payoff Increase if Direct = 0	0.152	0.075	0.027	0.031
	(0.130)	(0.136)	(0.152)	(0.132)
Observations	53	53	85	85

Notes: the dependent variable is a dummy that indicates whether i accepts the greedy offer; the table reports average marginal effects estimated from probit models; for the treatment interaction, we report the differential marginal effects of Payoff Increase when Direct changes from 0 to 1 (conditional marginal effects are reported in the lower part of the table); for the interaction of Belief with Direct, we report the differential marginal effect of Belief when Direct changes from 0 to 1; standard errors in parentheses; significance levels are: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Next, we compare the treatment effects on A-subjects' behavior across genders. Figure 8 shows, for females and males separately, the average frequency of choosing the default allocation by payoff treatment and method of play. When splitting the sample, the average frequency of the default allocation does not differ significantly across methods of play neither for females (MW: p-value 0.20, 75 observations) nor for males (MW: p-value 0.18, 101 observations). Moreover, with the strategy method the frequency of default allocations does not differ across payoff treatments neither for females (MW: p-value 0.27, 34 observations) nor for males (MW: p-value 0.91, 32 observations). Instead, Payoff Increase has a significant effect with the direct response method for both categories, but in opposite directions for females and males. In particular, the average frequency of default allocations for females is higher in D1 than in D2 (MW: p-value 0.05, 41 observations), while for males it is higher in D2 than in D1 (MW: p-value 0.09, 69 observations).



Females Males

Figure 8: A-subjects' behavior by gender

Thus, only when limiting the attention to the male sample, A-subjects' react to the payoff treatment as expected. This is confirmed by a probit estimate of the probability of choosing the default allocation, conducted separately for males and females, whose results are reported in Table 8 below. We adopt alternative specifications I and III used above for the whole sample of A-subjects (see Table 6). We find that the average marginal effect of Payoff Increase on the females' probability of choosing the default allocation is negative and significant, which is at odds with our predictions. Instead, the average marginal effect of Payoff Increase for males, conditioning on the direct response method, is positive and mildly significant, while, when conditioning on the strategy method, it is positive but not significant.

Table 8: A-subjects' behavior by gender

	Females		Males	
	I	II	I	II
Direct	0.142	0.063	-0.127	-0.139
	(0.111)	(0.110)	(0.097)	(0.089)
Payoff Increase	-0.253**	-0.228**	0.124	0.094
	(0.111)	(0.105)	(0.084)	(0.077)
Direct × Payoff Increase	-0.119	-0.029	0.153	0.158
	(0.223)	(0.215)	(0.193)	(0.175)
Belief		-0.006***		-0.007***
		(0.002)		(0.001)
Morality		0.004		0.024
		(0.100)		(0.072)
Guilt aversion		0.031		0.013
		(0.056)		(0.034)
Age		-0.047		-0.008
		(0.040)		(0.023)
Payoff Increase if Direct = 1	0.307**	-0.244*	0.172*	0.144*
	(0.150)	(0.146)	(0.095)	(0.087)
Payoff Increase if Direct = 0	-0.188	-0.215	0.020	-0.014
	(0.165)	(0.156)	(0.168)	(0.153)
Observations	75	75	101	101

Notes: the dependent variable is a dummy that indicates whether i chooses the default allocation; the table reports average marginal effects estimated from probit models; for the treatment interaction, we report the differential marginal effects of Payoff Increase when Direct changes from 0 to 1 (conditional marginal effects are reported in the lower part of the table); standard errors in parentheses; significance levels are: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Appendix 4

Experimental instructions

The following instructions apply for all treatments, unless otherwise specified.

Welcome and thank you for participating in this experiment. By closely following the instructions you will have the chance to earn an amount of money that will be paid in cash at the end of the experiment. You are not allowed to talk or communicate with other participants. If you have any questions, raise your hand and an assistant will help you.

Instructions

The experiment relates to the two-player game described below. At the beginning of the experiment you will be randomly assigned the role of **Player 1** or of **Player 2**. You will be paired randomly and anonymously with another participant with a different role. There will be **11 pairs of participants** in total.

The game

Player 1 can make **Offer A** or **Offer B** to Player 2. Every offer is a way to assign an amount of money to each player. If Player 1 makes Offer A, the two players obtain automatically the amounts of money that this offer assigns to them. If Player 1 makes Offer B, Player 2 can either **Accept** it or **Reject** it.

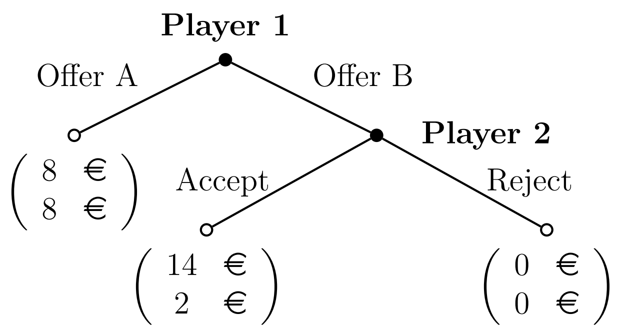
If Player 1 makes Offer A

- Player 1 obtains **8 EUR**.
- Player 2 obtains **8 EUR** (***D1 and S1***). [Player 2 obtains **11 EUR** (***D2 and S2***).]

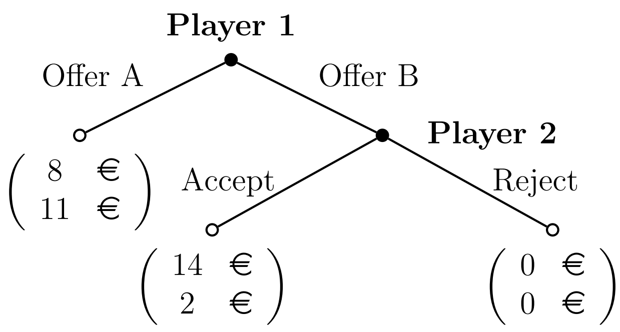
If Player 1 makes Offer B and Player 2 Accepts it

- Player 1 obtains **14 EUR**.
- Player 2 obtains **2 EUR**.

If Player 2 Rejects it, both players get **0 EUR**. The players' choices jointly determine the outcome of the game and the players' earnings as illustrated in the following Figure. Note that the first number in parenthesis indicates Player 1's earnings, while the second indicates Player 2's earnings.



(a) Figure in the instructions for D1 and S1



(b) Figure in the instructions for D2 and S2

How to make a choice (*D1 and D2*)

Player 1 decides first and chooses whether to make Offer A or Offer B, then his/her choice is communicated to Player 2.

Once informed about Player 1's choice, Player 2 can find himself in either one of the following situations:

- if Player 1 made Offer A, Player 2 does not have to make any choice and the players obtain the amounts of money assigned to them by Offer A.
- if Player 2 made Offer B, Player 2 has to decide whether to accept or reject it and his decision determines the players' earnings as described above.

[**How to make a choice (*S1 and S2*)**

Player 1 decides whether to make Offer A or Offer B. Player 2 has to make a choice without knowing whether Player 1 has made Offer A or Offer B. Hence, he/she has to decide whether he/she would Accept or Reject Offer B in case he/she received it. Once both players have decided, their choices are combined and

- if Player 1 made Offer A, Player 2's choice is ignored and the players obtain the amounts of money assigned to them by Offer A.
- if Player 2 made Offer B, Player 2's choice applies and determines the players' earnings as described above.]

You will now answer some control questions that will appear on the screen to verify whether you understood the instructions so far.

The computer has formed the 11 pairs and informed you about your role in the pair. With this role you will make a **CHOICE** in the game.

Before making a choice, you will be asked to make a **PREDICTION** about the choices that will be made in the game by participants of the **other 10 pairs** playing in a role different from yours.

Through your prediction you will have the chance to earn money as explained below. Note that you will be paid either what you earned in the prediction task or in the game. At the end of the experiment the computer will randomly determine for each pair whether the two players will be paid their earnings from the prediction task (probability 50%) or from the game (probability 50%). Earnings obtained in the selected part plus 4 EUR for filling a questionnaire will be paid in cash to you at the end of the experiment.

Since the prediction tasks are different for the two players, we describe them separately.

How to make a PREDICTION: Player 1 (*D1 and D2*)

Consider participants in the role of Player 2 who have not been paired to you and have received Offer B. You will be asked to predict in which percentage they will **Accept Offer B**. You can express your prediction by indicating a percentage between 0% and 100% (rounding up to integer numbers). After all participants have made their choice in the game, you will be informed about the actual percentage of participants of the other pairs who accepted Offer B among those who have received it.

If the distance between the percentage you have predicted x% and the actual one y% is smaller than 10%, i.e., if

$$|x\% - y\%| < 10\%$$

you obtain **10 EUR**, otherwise you obtain **2 EUR**.

Note that you receive 2 EUR both in case you underestimate and in case you overestimate the actual percentage with a distance larger than 10 percentage points. Moreover, you obtain 2 EUR even if the distance is exactly equal to 10 percentage points.

In case none of the 10 participants have received Offer B, the computer will determine randomly (with probability 50%) whether you obtain 10 or 2 EUR.

[**How to make a prediction: Player 1 (*S1 and S2*)**

Consider the participants in the role of Player 2 who have not been paired to you.

You will be asked to predict how many of them will **Accept Offer B** in case they received it.

You can express your prediction by indicating a number between 0 and 10.

After all participants have made their choice in the game, you will be informed about the actual number of participants of the other pairs who have decided that they would Accept Offer B in case they received it. If the number you indicated coincides with the actual number you obtain **10 EUR**, otherwise you obtain **2 EUR**.]

How to make a PREDICTION: Player 2

Consider the participants in the role of Player 1 who have not been paired to you. You will be asked to predict in which percentage they will make **Offer A**.

You can express your prediction by indicating a number between 0 and 10.

After all participants have made their choice in the game, you will be informed about how many participants of the other pairs made Offer A. If the number you indicated coincides with the actual number you obtain **10 EUR**, otherwise you obtain **2 EUR**.

Summing up the timing of the experiment is as follows:

- Predictions;
- Choices in the game;
- You are informed about the earnings (both yours and of the other player) from the predictions and from the game;
- The computer selects either the game or the predictions for the payment and informs you about your final earnings;
- Questionnaire;
- Payment.

You will now answer some control questions that will appear on the screen to verify whether you understood the instructions.

Acknowledgements

We would like to thank Giuseppe Attanasi, Federico Bobbio, Lorenzo Cappellari, Anna Conte, Roberto Corrao, Enzo Di Pasquale, Martin Dufwenberg, Antonio Filippin, Werner Güth, Vittoria Levati, Emanuele Lo Gerfo, Paola Moscariello, Giulio Principi, Simona Sacchi, Dov Samet, Alec Smith and seminar participants at the Second Workshop on Psychological Game Theory held at the University of East Anglia and at the BEEN Workshop held at the University of Bologna for useful comments. Pierpaolo Battigalli gratefully acknowledges financial support from Bocconi University and ERC grant 324219.

References

[1] Andersen S, Ertaç S, Gneezy U, Hoffman, List JA (2011) Stakes matter in Ultimatum Games. *American Economic Review* 101:3427-3439.

[2] Andreoni J, Vesterlund L (2001) Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics* 116:293-312.

[3] Aquino K, Reed A (2002) The self importance of moral identity. *Journal of Personality and Social Psychology* 83:1423-1440.

[4] Attanasi G, Battigalli P, Manzoni E (2016) Incomplete information models of guilt aversion in the Trust Game. *Management Science* 62:648-667.

[5] Attanasi G, Battigalli P, Nagel R (2013) Disclosure of belief-dependent preferences in a Trust Game. IGIER Working Paper 506, Bocconi University, Milan.

[6] Azar OH, Lahav Y, Voslinsky A (2015) Beliefs and social behavior in a multiperiod ultimatum game. *Frontiers in Behavioral Neuroscience* 9:29.

[7] Bar-Hillel M, Budescu D (1995) The elusive wishful thinking effect. *Thinking and Reasoning* 1:71-103.

[8] Battigalli P, Dufwenberg M (2009) Dynamic psychological games. *Journal of Economic Theory* 144:1-35.

[9] Battigalli P, Charness G, Dufwenberg M (2013) Deception: The role of guilt. *Journal of Economic Behavior and Organization* 93:227-232.

[10] Battigalli P, Corrao R, Dufwenberg M (2019a) Incorporating belief-dependent motivation in games. *Journal of Economic Behavior and Organization* 167:185-218.

[11] Battigalli P, Dufwenberg M, Smith A (2019b) Frustration, aggression, and anger in leader-follower games. *Games and Economic Behavior* 117:15-39.

[12] Battigalli P, Gilli M, Molinari C (1992) Learning and convergence to equilibrium in repeated strategic interaction. *Research in Economics* 46:335-378.

[13] Bauer D, Wolff I (2018) Biases in beliefs: Experimental evidence. TWI Research Paper Series, 109.

[14] Binmore KG, Gale J, Samuelson L (1995) Learning to be imperfect: The ultimatum game. *Games and Economic Behavior* 8:56-90.

[15] Blount (1995) When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes* 63:131-144.

[16] Blount S, Bazerman MH (1996) The inconsistent evaluation of absolute versus comparative payoffs in labor supply and bargaining. *Journal of Economic Behavior and Organization* 30:227-240.

[17] Bohnet I, Zeckhauser R (2004) Social comparisons in ultimatum bargaining. *Scandinavian Journal of Economics* 106:495-510.

[18] Bolton GE, Zwick R (1995) Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior* 10:95-121.

[19] Bosman R, van Winden F (2002) Emotional hazard in a Power-to-Take experiment. *The Economic Journal* 112:147-169.

[20] Bosman R, Sutter M, van Winden F (2005) The impact of real effort and emotions in the Power-to-Take game. *Journal of Economic Psychology* 26:407-429.

[21] Brandts J, Charness G (2011) The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics* 14:375-398.

[22] Brosig J, Weimann J (2003) The hot versus cold effect in a simple bargaining experiment. *Experimental Economics* 6:75-90.

[23] Camerer C, Lovallo D (1999) Overconfidence and excess entry: An experimental approach. *American Economic Review* 89:306-318.

[24] Cameron LA (1999) Raising the stakes in the Ultimatum Game: Experimental evidence from Indonesia. *Economic Inquiry* 37:47-59.

[25] Caplin A, Leahy J (2001) Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of Economics* 116:55-79.

[26] Card D, Dahl GB (2011) Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics* 126:102-143.

[27] Carpenter JP, Verhoogen E, Burks S (2005) The effect of stakes in distribution experiments. *Economics Letters* 86:393-398.

[28] Chang LJ, Sanfey AG (2013) Great expectations: Neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience* 8:277-284.

[29] Charness G, Rabin M (2002) Understanding social preferences with simple tests. *The Quarterly Journal of Economics* 117:817-869.

[30] Chen DL, Shonger M, Wickens C (2016) oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9:88-97.

[31] Civai C, Corradi-Dell'Acqua C, Gamer M, Rumiati RI (2010) Are irrational reactions to unfairness truly emotionally-driven? Dissociated behavioural and emotional responses in the Ultimatum Game task. *Cognition* 114:89-95.

[32] Cohen TR, Wolf ST, Panter AT, Insko CA (2011) Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology* 100:947-966.

[33] Costa-Gomes M, Weizsäcker G (2008) Stated beliefs and play in normal form games. *Review of Economic Studies* 75:729-762.

[34] Dufwenberg M, Kirchsteiger G (2004) A theory of sequential reciprocity. *Games and Economic Behavior* 47:268-298.

[35] Dufwenberg M, Kirchsteiger G (2014) Reciprocity and the ultimatum game. In: How Werner Güth's ultimatum game shaped our understanding of social behavior (van Damme E, Binmore KG, Roth AE, Samuelson L, Winter E, Bolton GE, Ockenfels A, Dufwenberg M, Kirchsteiger G, Gneezy U, Kocher MG) *Journal of Economic Behavior and Organization* 108:292-318.

[36] Dufwenberg M, Li F, Smith A (2019a) Promises and punishment. Unpublished manuscript.

[37] Dufwenberg M, Li F, Smith A (2019b) Threats. Unpublished manuscript.

[38] Eagly AH, Wood W (2011) Social role theory. *Handbook of Theories in Social Psychology* 2:458-476.

[39] Eckel CC, Grossman PJ (1998) Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal* 108:726-735.

[40] Eckel CC, Grossman PJ (2001) Chivalry and solidarity in Ultimatum Games. *Economic Inquiry* 39:171-188.

[41] Falk A, Fehr E, Fischbacher U (2003) On the nature of fair behavior. *Economic Inquiry* 41:20-26.

[42] Falk A, Fischbacher U (2006) A Theory of Reciprocity. *Games and Economic Behavior* 54:293-315.

[43] Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114:817-868.

[44] Frank RH (1987) If homo economicus could choose his own utility function, would he want one with a conscience? *American Economic Review* 77:593-604.

[45] Frank RH (1988) *Passions within Reason: The strategic Role of the Emotions.* New York: W.W. Norton.

[46] Fudenberg D, Levine D (1998) *The Theory of Learning in Games.* Cambridge MA: MIT Press.

[47] Gabay AS, Radua J, Kempton MJ, Mehta MA (2014) The Ultimatum Game and the brain: A meta-analysis of neuroimaging studies. *Neuroscience and Biobehavioral Reviews* 47:549-558.

[48] Gilam G, Abend R, Shani H, Ben-Zion Z, Hendler T (2018) The anger-infused Ultimatum Game: A reliable and valid paradigm to induce and assess anger. *Emotion* 19:84-96.

[49] Gneezy U, Imas A (2014) Materazzi effect and the strategic use of anger in competitive interactions. *Proceedings of the National Academy of Sciences* 111:1334-1337.

[50] Gospic K, Mohlin E, Fransson P, Petrovic P, Johannesson M, Ingvar M (2011) Limbic justice–amygdala involvement in immediate rejection in the Ultimatum Game. *PLoS Biology* 9:e1001054.

[51] Grimm V, Mengel F (2011) Let me sleep on it: Delay reduces rejection rates in Ultimatum Games. *Economics Letters* 111:113-115.

[52] Güth W, Schmittberger R, Schwarze B (1982) An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3:367-388.

[53] Güth W, Huck S, Müller W (2001) The relevance of equal splits in Ultimatum Games. *Games and Economic Behavior* 37:161-169.

[54] Güth W, Tietz R (1990) Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology* 11:417-449.

[55] Güth W, Schmidt C, Sutter M (2007) Bargaining outside the lab—a newspaper experiment of a three-person Ultimatum Game. *The Economic Journal* 117:449-469.

[56] Hoffman E, McCabe KA, Smith, VL (1996) On expectations and the monetary stakes in Ultimatum Games. *International Journal of Game Theory* 25:289-301.

[57] Ivanov A (2011) Attitudes to ambiguity in one-shot normal-form games: An experimental study. *Games and Economic Behavior* 71:366-94.

[58] Koszegi B, Rabin M (2009) Reference-dependent consumption plans. *American Economic Review* 99:909-936.

[59] Loewenstein G, Schkade D (1999) Wouldn't it be nice? Predicting future feelings. In: *Well-being: The Foundations of Hedonic Psychology* (Kahneman D, Diener E, Schwarz N) Russell Sage Foundation, New York, 85-105.

[60] Miller GA, Levin DN, Kozak MJ, Cook III EW, McLean Jr A, Lang PJ (1987) Individual differences in imagery and the psychophysiology of emotion. *Cognition and Emotion* 1:367-390.

[61] Munyo I, Rossi MA (2013) Frustration, euphoria, and violent crime. *Journal of Economic Behavior and Organization* 89:136-142.

[62] Nyarko Y, Shotter A (2002) An experimental study of belief learning using elicited beliefs. *Econometrica* 70:971-1005.

[63] Oechssler J, Roider A, Schmitz PW (2015) Cooling off in negotiations: Does it work? *Journal of Institutional and Theoretical Economics* 171:565-588.

[64] Oosterbeek H, Sloof R, Van De Kuilen G (2004) Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics* 7:171-188.

[65] Oxoby RJ, McLeish KN (2004) Sequential decision and strategy vector methods in ultimatum bargaining: Evidence on the strength of other-regarding behavior. *Economics Letters* 84:399-405.

[66] Persson E (2018) Testing the impact of frustration and anger when responsibility is low. *Journal of Economic Behavior and Organization* 145:435-448.

[67] Pillutla MM, Murnighan JK (1996) Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes* 68:208-224.

[68] Rabin M (1993) Incorporating fairness into game theory and economics. *American Economic Review* 83:1281-1302.

[69] Reuben E, van Winden F (2008) Social ties and coordination on negative reciprocity: The role of affect. *Journal of Public Economics* 92:34-53.

[70] Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the Ultimatum Game. *Science* 300:1755-1758.

[71] Sanfey AG (2009) Expectations and social decision-making: Biasing effects of prior knowledge on ultimatum responses. *Mind and Society* 8:93-107.

[72] Seip EC, Van Dijk WW, Rotteveel M (2014) Anger motivates costly punishment of unfair behavior. *Motivation and Emotion* 38:578-588.

[73] Slonim R, Roth AE (1998) Learning in high stakes Ultimatum Games: An experiment in the Slovak Republic. *Econometrica* 66:569-96.

[74] Solnick SJ (2001) Gender differences in the Ultimatum Game. *Economic Inquiry* 39:189-200.

[75] Spielberger CD (1999) Staxi-2: state-trait anger expression inventory-2; professional manual. *Psychological Assessment Resources*.

[76] Sutter M, Kocher M, Strauß S (2003) Bargaining under time pressure in an experimental Ultimatum Game. *Economics Letters* 81:341-347.

[77] Thaler RH (1988) Anomalies: The Ultimatum Game. *Journal of Economic Perspectives* 2:195-206.

[78] van Kleef GA, De Dreu CKW, Manstead ASR (2004) The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology* 86:57-76.

[79] van Leeuwen B, Noussair C, Offerman T, Suetens S, van Veelen M, van de Ven J (2017) Predictably angry facial cues provide a credible signal of destructive behavior. *Management Science* 64:3352-3364.

[80] van Wout M, Kahn RS, Sanfey AG, Aleman A (2006) Affective state and decision-making in the Ultimatum Game. *Experimental Brain Research* 169:564-568.

[81] Weinstein ND (1989) Optimistic biases about personal risks. *Science* 246:1232-1234.

[82] Xiang T, Lohrenz T, Montague PR (2013) Computational substrates of norms and their violations during social exchange. *The Journal of Neuroscience* 33:1099-1108.

[83] Xiao E, Houser D (2005) Emotion expression in human punishment behavior. *PNAS* 102:7398-7401.