

Lying about the price? Ultimatum bargaining with messages and imperfectly observed offers

Nejat Anbarci

School of Accounting, Economics and Finance
Deakin University
Burwood VIC 3125, Australia
nejat.anbarci@deakin.edu.au

Nick Feltovich*

Department of Economics
Monash University
Clayton VIC 3800, Australia
nicholas.feltovich@monash.edu

Mehmet Y. Gürdal

Department of Economics
Boğaziçi University
Bebek, Istanbul, Turkey TR–34342
mehmet.gurdal@boun.edu.tr

December 5, 2014

Abstract

We introduce the *taxicab game*, related to the ultimatum game and Gehrig et al.'s (2007) yes/no game. The proposer makes an *offer*, and simultaneously sends a cheap talk *message* indicating (possibly falsely) the amount of the offer. The responder observes the message with certainty and the offer with probability p before accepting or rejecting the offer. We investigate versions with $p = 0$ and $p = 0.5$ along with the ultimatum game as a baseline. Intuition and a model comprising both standard economic agents and others who dislike inequity, lies and lying provide clear predictions that our experimental results support. As the likelihood increases of offers being seen, the offers themselves increase, messages over-state them less, and responders are more likely to accept (even when the offer is unseen). Also, responders are more likely to accept after truthful messages than after lies or when no message is sent.

Journal of Economic Literature classifications: C72, C78, D82.

Keywords: ultimatum game, messages, lies, truth-telling, other-regarding behaviour.

*Corresponding author. Some of this research took place while Gürdal was at TOBB University of Economics and Technology. Financial support from Deakin and Monash Universities is gratefully acknowledged. We thank an editor, three anonymous referees, Sheryl Ball, Catherine Eckel, Nisvan Erkal, Philip J. Grossman, Kenan Kalayci, Ismail Saglam and Randy Silvers for helpful suggestions and comments.

1 Introduction

A traveller is about to hire a taxi at the airport in an unfamiliar city. While the cab's window displays aspects of the pricing scheme (fixed "flag fall" fee, rates per unit of distance and per minute of waiting time, etc.), the traveller does not know the actual amount she will be charged for a particular trip. She tells the cab driver the name of her hotel and asks how much the ride will cost. The driver quotes a price, after which the traveller must decide whether to hire the taxi.

This game between the traveller and taxi driver – as just described – is similar to the well-known *ultimatum game* (Güth, Schmittberger and Schwarze, 1982). However, there is an extra complication. If the traveller does choose to hire the cab, the price she eventually pays may not be the same as the price quoted by the driver. Both driver and traveller understand that the quote is non-binding, and in particular, the driver may well know that the actual price (based on the taxi-meter, but in effect set by the driver by his choice of route to the hotel) is going to be higher. The traveller will not learn the actual price until the end of her trip, and therefore must make her decision based only on the quote.

We call this the *taxicab game*, following our example. Like the ultimatum game, there is a fixed sum of money (a "cake") available to be split, with one player (the *proposer*) making a single offer to the other player (the *responder*). Unlike the ultimatum game, the proposer also sends a message to the responder, of the form "I have offered you X". The message is *cheap talk* in the game-theoretic sense (costless and non-binding) and in particular lies are quite possible. In the basic taxicab game, the responder is informed of the message but not the offer, and must therefore decide to accept or reject based only on the message. Thus it is very similar to Gehrig et al.'s (2007) *yes/no game* – the only difference being the message sent by the proposer to the responder in the taxicab game (in the yes/no game, the responder makes his decision based on no information at all). As in both the yes/no game and the ultimatum game, acceptance in the taxicab game means the responder receives the amount offered (the message is not directly payoff-relevant) and the proposer receives the remainder, while rejection means that a disagreement outcome is imposed. We also consider a *modified taxicab game* which is identical to the basic taxicab game except that the responder is able to observe the offer with a strictly positive probability (while still receiving the message with certainty). In Section 3.1 we discuss in more detail the relationships between our taxicab games, the yes/no game, and the ultimatum game.

Games like our taxicab game – where one player has private information about the proposed *distribution* of a surplus whose *size* is common knowledge – have been the subject of much less theoretical and experimental study than a related class of games where the asymmetric information is about the size of the surplus (with the yes/no game a notable exception, of course). They deserve attention in their own right, though, as they may serve as a model for many interactions where one side has an informational advantage regarding the ultimate transaction price of a good. This may be especially common in service markets. The game between a motorist whose car broke down in a remote area with only one nearby mechanic, and the mechanic who gives a quote for the required work that might be quite different from the eventual price charged, is another example. Still another is the "hidden cost" game played by a bank or utility company with its customers. Such information asymmetries can also arise in labour markets; for example, workers in a firm may have private information about how much employee theft (which may be thought of as a "top-up" of compensation beyond the amount contracted with the firm) they intend to carry out. And with a minor extension (to accommodate multiple proposers rather than one), the game can be applied readily to public-choice settings, such as the game played between political candidates – making campaign promises about the level of rents they will appropriate if elected – and the voters/taxpayers from whom these rents are appropriated

(see Corazzini et al., 2013, for an example).

This paper is an investigation of two versions of the taxicab game – the basic game and a modified game where the responder observes the offer with probability one-half (denoted TG(0) and TG(0.5) respectively) – using the ultimatum game (UG) as a baseline, and with disagreement payoffs equal to zero.¹ Under the usual *Homo economicus* assumption of pure own-money-payment maximisation, these are trivial to analyse theoretically: responders in all three games should ignore messages and accept any positive offer, and therefore proposers should offer either zero or the smallest monetary unit. However, we will relax the assumption of pure own-money-payment maximisation, in favour of a view that some people exhibit (1) “inequity aversion”, a distaste for either favourable or unfavourable inequality of money payoffs, and (2) “deception aversion”, a distaste for either lying or being lied to. A substantial body of experimental-economics research suggests that a non-trivial fraction of people have such tastes (see Section 2 for a short survey), and when the population comprises some of these “other-regarding” players as well as the usual “self-regarding” players, we obtain clear comparative-static predictions of how behaviour differs across games as well as across contingencies within a game. (In Section 3.3, we describe these predictions along with some informal intuition behind them, but for a more rigorous treatment, see Appendix A, where we develop and analyse a theoretical model of self- and other-regarding players.)

We test these predictions with a laboratory experiment. Behaviour in the baseline UG is broadly similar to ultimatum-game results in many other experiments, providing a useful point of reference for the other two treatments. Comparisons within and across treatments largely resemble the comparative statics of the model. First, proposers’ offers are lower in both taxicab games than in the UG, and lower still in TG(0) than TG(0.5); that is, offers become smaller as their likelihood of being seen decreases. Second, while lying (sending a message that over-states the corresponding offer) is commonplace, with the average message in both taxicab games 50–100 percent higher than the average offer, it is not universal (about one-fifth of messages are truthful, and another one-tenth actually under-state the offer), and its extent varies with the game: about 40 percent of messages are either truthful or under-statements in TG(0.5), versus about 20 percent in TG(0). Third, responders are significantly more likely to accept an unseen offer in TG(0.5) than in TG(0) – that is, when the offer was made knowing it might be observable versus when it could never have been – and in TG(0.5), unseen offers are more likely than seen ones to be accepted. Fourth, in those cases of TG(0.5) where responders observe the offer as well as the message, acceptance is substantially less likely when the message under-states the offer (i.e., the proposer is caught having lied about the offer); however, comparison to the UG (where no message is sent) suggests that responders might be rewarding truth-telling rather than punishing lying.

2 Other relevant research

Our study contributes to two related but distinct segments of the theory and experimental literature: (1) variants of the ultimatum game with incomplete information and/or cheap talk, and (2) lying in general strategic settings (not necessarily the UG) where cheap-talk messages are sent. Here, we describe some of the relevant literature.

Early experiments involving incomplete information and cheap talk in the ultimatum game tended to find subjects opportunistically taking advantage of their extra information, and using messages to deceive their opponents about

¹In the real game between traveller and taxi driver, payoffs following disagreement are likely not to be zero, since the traveller has alternative means of transport to the hotel, and the taxi driver will likely find other fares. The distinction between zero and positive disagreement payoffs matters, because when offers can be less than the responder’s disagreement payoff, even a self-regarding responder may strictly prefer to reject an unseen offer. Hence, our use of zero disagreement payoffs as a normalisation entails some loss of generality.

the nature of their information, but seldom to the extent predicted by standard game theory with self-regarding players.² A typical example is from one of the first papers we know of in this literature. Mitzkewitz and Nagel (1993) study a version of the ultimatum game where the cake size (between 1 and 6) is determined by a die roll and the realisation is told to the proposer but not the responder. In their “offer game”, the responder is informed of the proposer’s offer, while in their “demand game”, the responder is instead informed about the proposer’s demand (i.e., the cake size minus the offer). They find significant differences in behaviour between the two games, but it is the nature rather than the existence of these differences that is most interesting. In the offer game, proposers drawing a large cake size tried to pool with those drawing a smaller cake size: the modal offer for cake sizes of 1–4 is half of the cake, but for cake sizes of 5–6 is only 2. In the demand game, the reverse happened, with proposers most frequently demanding half of a cake size of 6, but 3 when the cake size is 3, 4 or 5, and the entire cake when it is 1 or 2.

Later studies have found mixed results. Kagel, Kim and Moser (1996) found no difference in offers according to whether proposers or responders are informed of the other’s exchange rate between tokens – which the subjects bargain over – and real money. Croson (1996) manipulated whether the responder knows the cake size and whether offers are made in money units or percents of the cake, and observed that in the money–unit treatment, proposers made significantly higher offers when responders knew the cake size than when they did not. Kriss et al. (2013) examine ultimatum games where only the proposer is informed about the cake size, and distinguish between implicit deception (e.g., making offers that would be “fair” if the cake had been smaller than it actually is) and explicit deception (sending a message containing a lie about the cake size). They find increased misrepresentation when explicit deception is possible compared to when only implicit deception is.

Gehrig et al. (2007) introduce the “yes/no game”, which as noted above, differs from our taxicab game only in not allowing messages. Gehrig et al. find that offers in the yes/no game are comparable to those in a dictator game (an ultimatum–game variant where the responder has no move, so the proposer’s proposal is automatically implemented), and substantially lower than in an ultimatum game. Interestingly, responders tend not to use their veto power, even though they were very likely to have received a small offer. Güth and Kirchkamp (2012) compare Gehrig et al.’s results to those from several field experiments with varying subject pools (e.g., business executives) and forms of interaction (playing by post or Internet rather than in the lab). Broadly, proposer behaviour is indistinguishable from the original lab experiment, but responders reject much more often in all of the new experiments.

Several studies focus on sender–receiver games, where players’ payoffs depend on the actions of the receiver, to whom the sender can transmit costless messages. Gneezy (2005) classifies lies into four types, depending on whether they help or harm the sender and whether they help or harm the receiver. He concentrates on one of these cases – where lying helps the sender at the expense of the receiver – and varies the size of these potential gains and losses. He finds that raising the sender’s benefit from lying, or reducing the harm imposed on the receiver, is associated with more lying. In Lundquist et al. (2009) subjects designated as sellers obtain scores on a general knowledge test and then send messages about their score to the buyer with whom they are matched. Buyers decide, upon receiving the message but without observing the score itself, whether to engage in a contract which is profitable if and only if the buyer’s actual score is above a commonly–known threshold. Lundquist et al. find substantial but not universal deception: depending on the treatment, 40 to 76 percent of subjects with scores below the cutoff over–state their scores. Interestingly, they also report evidence of an aversion to lying by sellers that increases in the “size” of the lie (i.e., the difference between the true and reported scores), even though the monetary cost to the buyer from such

²There is also an earlier literature on incomplete information and cheap talk in unstructured bargaining; see, for example, Roth and Murnighan (1982, 1983).

a lie is always the same. They also find that free-form communication implies a lower frequency of lies compared to pre-specified communication. López-Pérez and Spiegelman (2012) examine a setting where the informed sender cannot affect the uninformed receiver's payoffs. A random draw determines the colour of a circle (blue or green), and the sender sends the receiver a message indicating the colour. The sender receives a payoff of 15 by sending a green message, versus 14 from sending blue, while the receiver gets a constant 10. Even though lying does not harm the receiver, nearly half of senders report truthfully when the circle is blue.

The common result that subjects lie less often than predicted by standard theory has led some researchers to propose theoretical models assuming that individuals incur some cost to lying, but otherwise act according to standard game theory. Many of these models consider a related, but distinct, concept of honesty to ours. Rather than having a taste for correctly reporting a piece of private information (e.g., the cake size, or one's offer), these models consider honesty to be a taste for honouring past agreements or promises about one's own subsequent behaviour.³ This can be considered a cost of lying – in the event that a player entered an agreement while intending to violate it – but also means that players can incur the cost even when they entered the agreement with the intention of honouring it (e.g., if it was only afterward, upon further introspection or the receipt of new information, that the player decided not to honour the agreement).

Ellingsen and Johannesson (2004) examine this form of honesty using a game between a buyer and a seller, where the seller acts first and decides whether to pay a fixed cost to generate a surplus. After surplus generation, they play an ultimatum game with the buyer as proposer (so that after rejection, the seller loses the fixed cost). They consider a “promises” treatment where buyers can send pre-play messages, a “threats” treatment where sellers can send messages along with their decision, and a baseline with no messages. They report that either kind of communication increases offers – from an average of about 50 percent of the cake in the baseline to about 63 percent under seller messages and 70 percent under buyer messages (an offer of 60 percent is needed to cover the seller's fixed cost) – and this is anticipated by sellers, who are more likely to incur the fixed cost under either type of communication. Ellingsen and Johannesson develop a theoretical model combining Fehr and Schmidt's (1999) inequity aversion and a lump-sum disutility of lying, and use it to explain their main experimental results.⁴

López-Pérez (2012) introduces a more complex theoretical model with three types of player: self-regarding types, “H” players who incur a disutility by violating norms of honesty (proportional to the strength of the norm), and “EH” players who are motivated by a norm of equity/efficiency as well as the honesty norm. He shows that this model can explain many patterns of behaviour found in social-dilemma experiments with pre-play communication, and that an analogous model with H types but no EH types could not do so. Miettinen (2013) studies a model where the cost of reneging on pre-play agreements is increasing in the level of harm inflicted on others, and shows that the extent to which such “guilt” improves welfare depends on the game's structure. Games with strategic substitutes give rise to a conflict between efficiency and incentives to honour the agreement; this conflict does not arise when the strategies are strategic complements.

The closest theoretical model to the one we use (in Appendix A) is the one proposed by Besancenot et al. (2013) for the UG with asymmetric information about the cake size. Like Ellingsen and Johannesson (2004) and our model, they posit aversion to disadvantageous inequity (though not to advantageous inequity as we do) and a disutility of lying, though unlike Ellingsen and Johannesson, their disutility-of-lying function is linear in the size of the lie, and

³As evidence of costs of lying in our sense of mis-reporting private information, Abeler, Becker and Falk (2014) find that subject report the results of coin tosses approximately honestly on average, even with monetary incentives to lie, while Fischbacher and Föllmi-Heusi (2013) find a mix of honest subjects, complete liars (money-maximisers) and partial liars in reports of die rolls.

⁴See also Chen, Kartik and Sobel (2008) and Kartik (2009). Also, Charness and Dufwenberg (2006, 2010) make progress toward disentangling various explanations for individuals forgoing monetary gains in order to tell the truth or to punish liars.

unlike our model, they assume no disutility for the responder of being lied to. Besancenot et al. show theoretically that the presence of some truth-tellers makes it worthwhile for other proposers to lie, and in an experiment verify that lying is rampant (nearly 90 percent of proposers under-state the cake size).

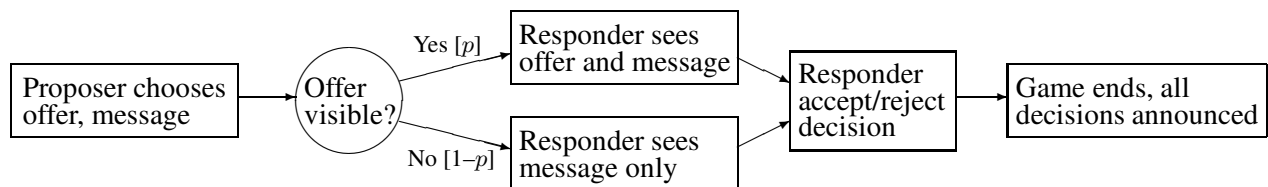
The work discussed above has tended to concentrate on whether message senders will lie when given the opportunity. However, there is also a literature looking at message receivers' response to being told lies. Sánchez-Pagés and Vorsatz (2007) consider a sender-receiver game with multiple possible states of the world, where only the sender knows the true state. They observe that senders send truthful messages about the state significantly more often than the theoretical prediction.⁵ Notably, though, they also find that receivers will expend resources to punish lying senders when this option is available, suggesting an aversion to being lied to. Peeters, Vorsatz and Walzl (2008) show, using similar games, that a non-negligible fraction of receivers will choose to *reward* senders when the opportunity is available, suggesting that some people receive a benefit from being told the truth. However, these rewards are found to have no significant effect on the overall truth-telling levels of senders (indeed, truth-telling seems to decrease). Brandts and Charness (2003) find evidence of both rewarding truth-telling and punishing lies in a 2x2 game preceded by cheap talk.

Finally, while our notion of deception aversion involves lies specifically, it is related to several other notions of disutility from bad treatment more generally. A disutility of lying is in the spirit of "guilt aversion" (Charness and Dufwenberg, 2006; Ellingsen et al., 2010) or "letting-down aversion" (Dufwenberg and Gneezy, 2000), while a disutility from being lied to is similar to "betrayal aversion" (Bohnet and Zeckhauser, 2004; Aimone and Houser, 2012; Aimone, Ball and King-Casas, 2013). Punishing lying in the taxicab game is akin to punishing breaches of trust in games like the trust game (Ohtsubo et al., 2010, show that even third parties are willing to punish such breaches, more so than unequal divisions that were obtained without lying).

3 The games and theoretical predictions

The experiment uses two versions of the taxicab game (TG) and, as a baseline, the standard ultimatum game (UG). In all three games, the two players bargain over a fixed sum of money ("cake"); this sum will be 20 Turkish liras (TRY) in the experiment, so for convenience we will set it to 20 here as well. In the UG, the proposer makes an offer $x \in \{0, 1, 2, \dots, 20\}$ to the responder. The responder sees x (with certainty) and chooses a response: Accept or Reject. If the responder accepts, monetary payoffs are $20 - x$ for the proposer and x for the responder; after a rejection, each receives a disagreement payoff of zero. In either case, the game ends with no opportunity for subsequent renegotiation. The rules of the game, including the cake size, the disagreement payoff, and the responder's information, are assumed to be common knowledge.

Figure 1: Sequence of events in the general taxicab game $TG(p)$



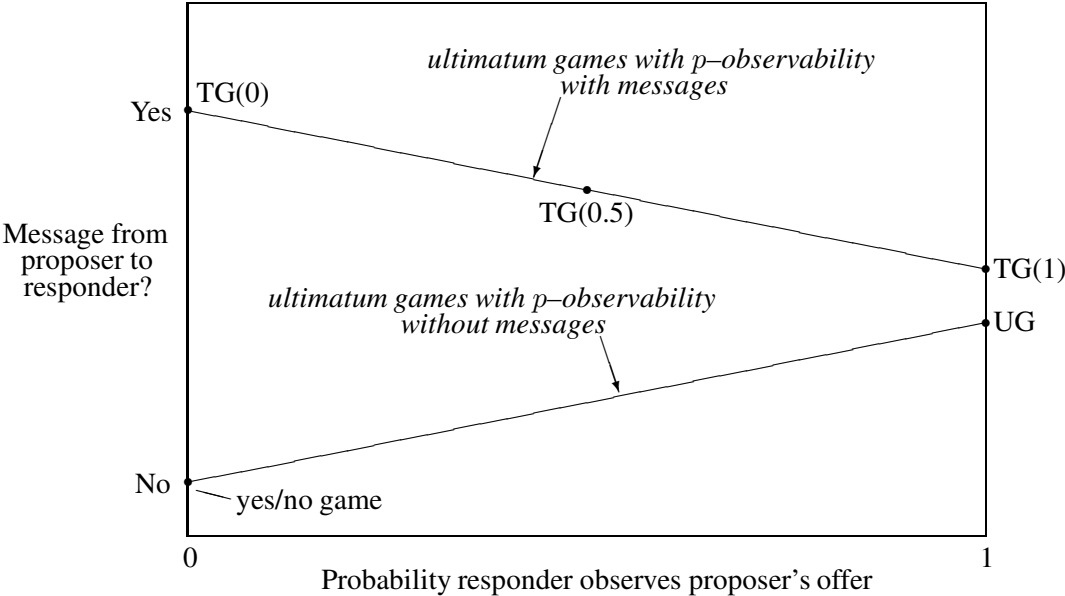
⁵Sánchez-Pagés and Vorsatz (2009) argue that this behaviour is consistent with aversion to lying rather than preference for truth-telling.

In the *basic taxicab game* $TG(0)$, the proposer sends a message $m \in \{0, 1, 2, \dots, 20\}$ in addition to the offer x . The message and offer spaces are identical by design; messages are meant to indicate the amount offered, and in the experiment are framed explicitly in this way. The responder observes m but not x , and then chooses Accept or Reject, after which monetary payoffs are determined as in the UG; in particular, m has no direct effect on either player’s monetary payoff. The *modified taxicab game* $TG(0.5)$ is similar, except that the responder *does* see the offer x with probability 0.5 (and still sees the message m with certainty) before choosing a response. Importantly, the proposer chooses x and m before knowing whether x will be seen by the responder. The time–line for the general version of the taxicab game $TG(p)$ (where p is the probability the responder sees the offer) is shown in Figure 1.

3.1 A partial taxonomy of games related to the taxicab game

Our taxicab game $TG(0)$ and modified taxicab game $TG(0.5)$ are members of a class of games called *ultimatum games with p –observability and with or without messages*, displayed in Figure 2. In all of these games, the proposer makes a single offer to the responder, after which the responder either accepts (in which case he gets the amount offered and the proposer gets the remainder) or rejects (in which case the disagreement outcome is imposed). These games differ in two dimensions. First, they differ in p , the probability that the responder sees the proposer’s offer: the probability is zero in the yes/no game and in our basic taxicab game; it is one in the UG; and it is one–half in our $TG(0.5)$ game. Second, they differ in whether the proposer sends a message about the offer to the responder: she does this in both the basic and the modified taxicab games, and does not in the UG and yes/no game.

Figure 2: Ultimatum games with p –observability and with/without messages



Obviously, other combinations besides the four discussed here are possible. The set of such combinations is represented by the two diagonal line segments in the figure. As it illustrates, the UG and yes/no game are special cases of ultimatum games with p –observability and *without* messages, while the basic and modified taxicab games are special cases of ultimatum games with p –observability and *with* messages. In particular, the game $TG(1)$ – with both the offer and message seen with certainty by the responder – is just the UG with an extra message. Clearly, since

the responder in TG(1) also knows the offer, the message ought to be superfluous. This is the reason for diagonal line segments in the figure rather than horizontal segments. When $p = 1$, the message is likely to have little value, implying that the UG and TG(1) are strategically very similar.⁶ But as p decreases, the responder is less likely to know the offer when choosing her response, so he must increasingly rely on the message. Of course, the message – being cheap talk – may be uninformative. In the next sections, we will theoretically analyse the TG(0) and TG(0.5), using the UG as a baseline, and we will see that while messages can be uninformative in the standard case where players care only about their own monetary payoffs, they can carry non-trivial information about the offer when this assumption is relaxed.

3.2 Predictions: self-regarding players

We will use the term “self-regarding” to describe players of either role (proposer or responder) whose objective is to maximise their own expected monetary payoff. If both players are self-regarding, solving these games is simple. Since $x \geq 0$ in all three games, in TG(0) and TG(0.5) self-regarding responders weakly prefer to accept even if the offer is unobserved, and irrespective of the message. Given this, backward induction implies that a self-regarding proposer will offer either 0 or 1 (the smallest positive allowable offer). Messages are either non-existent (in the UG) or – being only cheap talk – irrelevant to equilibrium (in TG(0) and TG(0.5)). Thus, if all players are self-regarding, there will be no meaningful differences in behaviour across the three games.

3.3 Predictions: other-regarding players

However, there is plenty of evidence that many people are not completely self-regarding, as discussed in Section 2. We will depart from the assumption of own-money-payoff maximisation in two particular ways. First, some individuals will be *inequity averse*: they dislike both advantageous inequity (getting a higher payoff than others) and disadvantageous inequity (getting a lower payoff than others). Second, some individuals will be *deception averse*: for proposers, this is an aversion to lying, while for responders it is an aversion to being lied to. We will define a message m to be a lie if it *over-states* the offer x ($m > x$), and truthful if $m \leq x$.⁷

We present a list of hypotheses for our experiment, along with some intuition behind them. In Appendix A, we derive these hypotheses from a simple theoretical model based on a population made up of both self- and other-regarding players. The model yields point predictions for behavioural variables such as offers, messages, and accept/reject decisions (see Table 9 in the appendix). However, since our focus is not on testing this particular model, we use these point predictions only for their associated comparative statics, which give rise to our hypotheses.

When some responders are other-regarding, proposers should understand that very low offers might be rejected if they are seen, so the more likely the proposer’s offer is to be seen, the more generous it should be.

⁶Since UG and TG(1) are so similar, it is likely that had our experiment included TG(1) in its design instead of UG, the results would not change substantially from what we observed, though we admit this is only conjecture. Our main reason for using the UG instead of TG(1) was to provide a baseline, using a well-known game, which would allow (depending on our results) either reassurance that our subject pool, experimental procedures, etc. are typical of those used in other lab experiments, or a way of controlling for any systematic differences when examining the TG(0) and TG(0.5) data. In the event, subjects in our UG behave fairly typically (see Section 5), and hence any differences between behaviour in our other treatments and standard ultimatum-game results are likely attributable to differences in the games’ strategic structures. We acknowledge that including a TG(1) treatment would have improved the direct comparability with the other treatments, and would also have allowed an examination of what messages proposers send when they know their offer will be observed with certainty.

⁷This abuses terminology to some extent (messages that under-state the offer are, in a literal sense, just as untrue as over-statements), but it captures the common moral distinction between lies that benefit the proposer at the responder’s expense and other kinds of untruth, which are typically viewed as either more benign or irrelevant (see, e.g., Gneezy, 2005).

Hypothesis 1 *Offers are higher in UG than TG(0.5), and higher in TG(0.5) than TG(0).*

Self-regarding responders should accept all positive offers, but other-regarding responders may reject sufficiently low offers, and the threshold offer for acceptance will depend on the strength of the responder's distaste for disadvantageous inequity, which may be heterogeneous across responders. Thus higher offers are more likely to be accepted than lower offers, so the treatment differences in offers according to Hypothesis 1 should lead to similar differences in acceptances.

Hypothesis 2 *Overall acceptance rates are higher in UG than TG(0.5), and higher in TG(0.5) than TG(0).*

Deception-averse responders who are able to observe both the offer and message will know whether they have been lied to, and for some offers, will reject if the message was a lie but accept if it was truthful.

Hypothesis 3 *In TG(0.5) when responders do observe the offer x , they are more likely to accept x when the message m is truthful.*

Deception-averse proposers may tell the truth even in TG(0) – where they know they will not be caught in a lie – while self-regarding proposers can lie with impunity there. In TG(0.5), however, Hypothesis 3 means that responders have some ability to punish lies, thus disciplining even self-regarding proposers to some extent, though not completely.

Hypothesis 4 *In both TG(0) and TG(0.5), messages are higher than offers.*

Hypothesis 5 *Messages are higher in TG(0) than TG(0.5).*

In TG(0), offers should be lower than in TG(0.5) (Hypothesis 1), while messages should be higher (Hypothesis 5). Thus, *over-statement* (the extent to which the message is higher than the corresponding offer) should be higher in TG(0).

Hypothesis 6 *The amount that messages over-state offers is higher in TG(0) than TG(0.5).*

Finally, since offers are higher in TG(0.5) than in TG(0), an inequity-averse responder who cannot see an offer should be more likely to accept it in TG(0.5).

Hypothesis 7 *Message-offer pairs in which the responder does not observe the offer are more likely to be accepted in TG(0.5) than TG(0).*

4 Experimental procedures

Each experimental session involved ten subjects playing one game (UG, TG(0) or TG(0.5)) for five rounds, followed by a questionnaire containing attitudinal and demographic questions.⁸ Each subject was randomly assigned a role – proposer or responder – that was fixed for the entire session, and was matched once to each of the five subjects of the opposite role.

⁸English translations of the instructions (which, along with the other materials, were written in Turkish) for the TG(0.5) treatment are in Appendix B, translated questions from the post-experiment questionnaire are in Appendix C, and sample screen-shots (also translated) are in Appendix D. Other experimental materials such as instructions from the other treatments, the raw data and computer programs are available from the corresponding author upon request.

All sessions were conducted by the same experimenter, and took place at the Social Sciences Laboratory at TOBB University of Economics and Technology in Ankara, Turkey in January 2013. Subjects, primarily undergraduate students, were invited by a school-wide email and registered online for a session; no one took part more than once. The experiment was run on networked personal computers, and programmed using the z-Tree experiment software package (Fischbacher, 2007). Subjects sat in individual carrels in a single room and were visually isolated from each other, and were asked to turn off their mobile phones and not to communicate with each other except via the computer program.

At the beginning of a session, subjects were given written instructions, which were also read aloud in an attempt to make them common knowledge. There was no instructions quiz, but subjects were given a chance to ask questions at this time (and throughout the session) which would be answered privately, after which the first round began. Each round began with a reminder of the subject's role. After that, the proposer was prompted to choose an offer, which could be any whole number of Turkish liras (TRY) between 0 and 20.⁹ In TG(0) and TG(0.5) the proposer decided at the same time what message she would send to the responder. This message was elicited with the phrasing, "In your message to player A, how many out of 20 Turkish liras will you tell him/her that you sent to him/her". The responder observed the offer in UG, the message in TG(0), and – with equal probability – either only the message or both the message and the offer in TG(0.5).¹⁰ He was then prompted to accept or reject the proposer's offer, after which payoffs were determined and subjects received feedback: the offer (irrespective of whether it was observed by the responder prior to his decision), the message when applicable, the responder's decision and the subject's payoff. Subjects were not told the opponent payoff, but received enough information to make that calculation if desired.

At the end of the last round, subjects completed the questionnaire, were paid privately and individually in cash, and left (no other experiment was conducted during the same session). For each session, one round was randomly chosen for payment, and subjects additionally received a show-up fee of 8 TRY. Total earnings averaged just under 16 TRY for a session that typically lasted about 25–30 minutes.

5 Experimental results

There were seventeen sessions, 5 each of the UG and TG(0) treatments, and 7 of the TG(0.5) treatment, for a total of 170 subjects. (We had more TG(0.5) sessions in order partly to offset the fact that in each session, half of observations have a seen offer and half have an unseen offer.) In order to facilitate comparison of our results with other experiments, we express all monetary quantities (e.g., offers) as percents of the cake size unless stated otherwise.

⁹At the time of the experiment, 1 TRY corresponded to USD 0.56 at market exchange rates. However, the lower cost of living in Turkey compared to many developed countries made the stakes correspondingly higher. For comparison, the minimum hourly wage in Turkey is 4.5 TRY, and a lunch at the school cafeteria costs about 6 TRY.

¹⁰Note that although the responder may not observe the proposer's offer until after making his decision, the offer is always observable to the *experimenter*. If the proposer cares about her image to the experimenter (e.g., if she gets disutility from knowing the experimenter knows she has lied), this may increase the cost of lying relative to the hypothetical case where no experimenter existed, or where a double-blind design meant that subjects were anonymous to the experimenter, though even then it is unlikely that truth telling would completely disappear (see Fischbacher and Föllmi-Heusi, 2013 for evidence toward this claim). Note also that the effect of observability to the experimenter ought to operate similarly in our TG(0) and TG(0.5) treatments.

5.1 Aggregate descriptive statistics

Table 1 shows some aspects of aggregate proposer behaviour in the experiment – offers, messages and over-statements (message minus offer, censored at zero) averaged over all rounds – along with results of non-parametric significance tests.¹¹ Figure 3 shows cumulative distributions of offers and messages for each treatment. Offers in

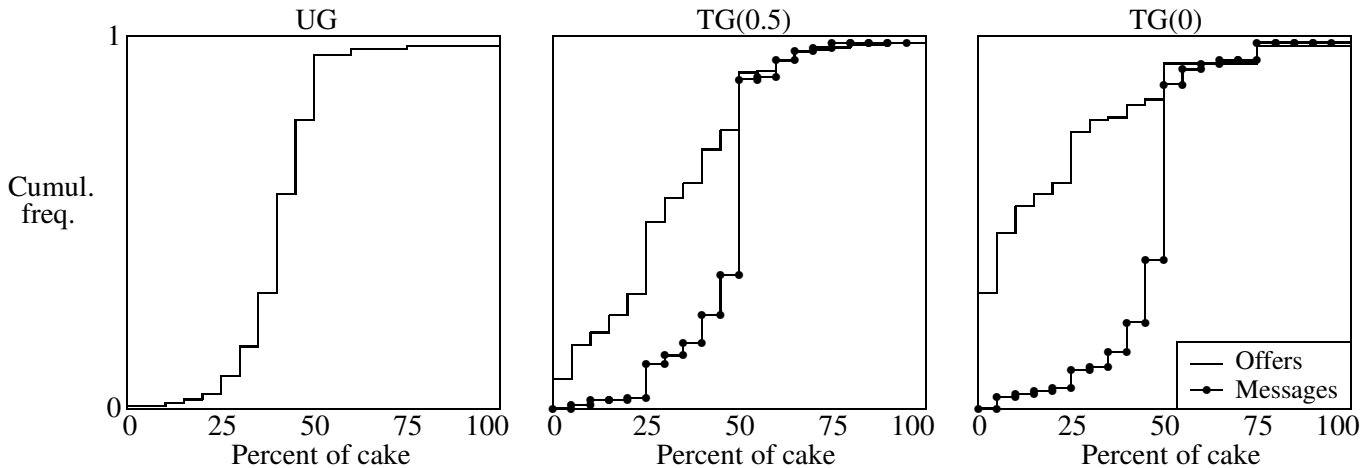
Table 1: Aggregate proposer data

	UG	TG(0.5)	TG(0)	Significantly different?
Mean offer (% of cake)	41.5	31.5	20.1	$p < 0.005$
Mean message (% of cake)	—	46.5	46.6	n.s.
Significantly different?		$p \approx 0.008$	$p \approx 0.03$	
Mean over-statement (% of cake)		17.7	29.8	$p \approx 0.01$

See text for descriptions of significance tests. n.s. = not significant ($p > 0.20$).

our UG treatment are fairly typical for ultimatum games; nearly all are between 20 and 50 percent of the cake, and they averaged a bit over 40 percent.

Figure 3: Observed distributions of offers and messages by treatment (cumulative frequencies)



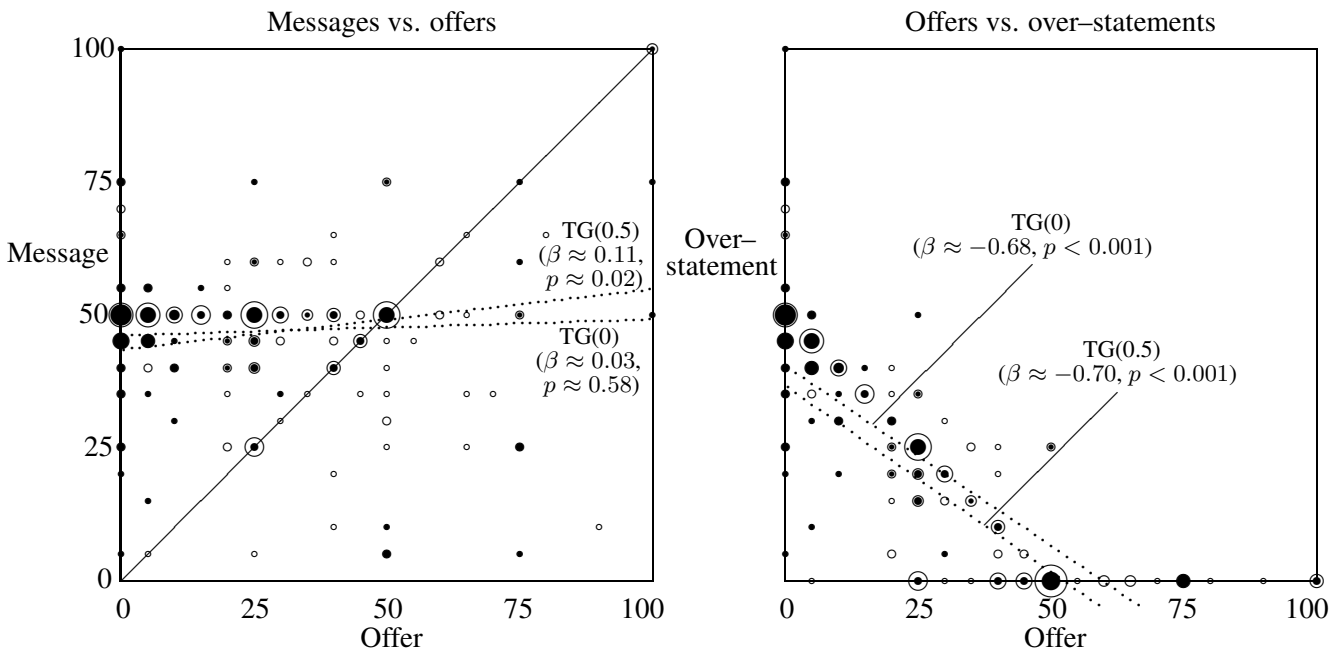
Consistent with Hypothesis 1, offers in both taxicab games are lower than those in the ultimatum game, with offers in TG(0) even lower than those in TG(0.5). There is substantial variance in offers in both TG treatments, with positive but small mass on the equal split. Despite this variance, the differences in offers are highly significant: the null hypothesis of no difference across the three treatments is rejected in favour of the alternative hypothesis of highest offers in UG and lowest offers in TG(0) (Kruskal–Wallis test, $p \approx 0.009$). Further support for treatment effects comes from pairwise robust rank–order tests, which find significantly higher offers in UG than in TG(0.5) ($p \approx 0.005$) and TG(0) ($p \approx 0.004$), and significantly higher offers in TG(0.5) than in TG(0) ($p \approx 0.03$).

¹¹All non-parametric statistical tests in this paper use session-level data, and all use two-tailed rejection regions unless stated otherwise. See Siegel and Castellan (1988) for descriptions of the tests we use, and see Feltovich (2005) for a table of critical values for the robust rank–order test.

We also see support for our hypotheses in the relationships between offers and messages. Consistent with Hypothesis 4, we see in Table 1 and Figure 3 that messages, the vast majority of which are close to 50 percent of the cake, are higher than offers in both TG(0.5) and TG(0), and these differences are significant (Wilcoxon signed-ranks tests, $p \approx 0.008$ for TG(0.5), $p \approx 0.031$ for TG(0)). Also, consistent with Hypothesis 6, the amount that messages over-state offers (i.e., the extent to which proposers lie on average) is significantly higher in TG(0) than in TG(0.5) (robust rank-order test, $p \approx 0.01$). However, we find no significant difference between treatments in the messages themselves, in contrast to Hypothesis 5’s prediction of higher-valued messages in TG(0) (robust rank-order test, $p > 0.20$).

Figure 4 shows every individual proposer (offer, message) pair in the taxicab game treatments in the left panel, and every (offer, over-statement) pair in the right panel. For each combination, the figure shows the number of times that combination was sent by proposers in the TG(0) treatment as the area of the dark circle, and the number of times it was sent in the TG(0.5) treatment as the area of the light circle (or ring surrounding a dark circle). Given

Figure 4: Proposer messages, offers, and over-statements as % of cake – TG(0) and TG(0.5) treatments (*The area of each circle is proportional to the number of observations with that (offer, message) or (offer, over-statement) pair. Within each circle, the dark circle has area proportional to the number of observations in the TG(0) treatment, and the light circle/ring has area proportional to the number of observations in the TG(0.5) treatment. Dotted lines represent OLS trends.*)



the aggregate descriptive statistics already seen, it is not surprising that the modal message is clearly 50 percent of the cake, or that there is little apparent correlation between offers and messages in either TG treatment. (Hence over-statement tends to increase as offers decrease, as shown in the right panel.) However, it is arguably noteworthy that there is any correlation at all between offers and messages – and thus that messages are informative to some extent. The figure shows OLS trend lines for the two treatments separately; both are positively sloped (though only the TG(0.5) slope is significantly different from zero).

Another surprising feature of the figure is the frequency of truthful messages. Proposers send a message exactly equal to their offer 13% of the time in the TG(0) treatment and 27% of the time in the TG(0.5) treatment. Further, a

non-negligible fraction of incorrect messages are in the “wrong” direction, with the proposer’s offer *higher* than the corresponding message 8% of the time in the TG(0) treatment and 13% of the time in the TG(0.5) treatment.¹² The difference in frequency of (strict) truth-telling between TG(0) and TG(0.5) is weakly significant (robust rank-order test, $p \approx 0.09$), while the difference in messages under-stating the offer is not significant ($p > 0.20$).

Table 2 reports results on proposer behaviour from four pairs of panel Tobit regressions.¹³ Models 1–4 have the offer as the dependent variable, using either all proposer data (Models 1–2) or the proposer data from the TG(0) and TG(0.5) treatments (Models 3–4). Models 5 and 6 also use the proposer data from the two TG treatments, but have the message as the dependent variable, while Models 7 and 8 use over-statement. Within each pair of models, there is a restricted model with only the main treatment variables, and an unrestricted model with additional explanatory variables. The main treatment variables in Models 1 and 2 are indicators for two of the three treatments (with UG as the baseline), the round number, and its product with the two treatment indicators. Models 3 and 4 use the TG(0.5) indicator (so that TG(0) is the baseline), the round number, the message, and all products of these variables (including the three-way product); Models 5–8 use the TG(0.5) indicator, the round number, and their product. The additional variables used in the unrestricted models (2, 4, 6 and 8) are an indicator for female; the subject’s age (to the nearest year); indicators for living with family, with friends, or alone (living in a dorm is the baseline); indicators for economics student and business (non-economics) student; number of younger siblings; number of older siblings; indicator for being an only child; number of economics classes completed (up to a maximum of 4); the 15 attitudinal variables from the questionnaire (see Appendix C); and the subject’s decision time (i.e., the time from the beginning of the stage until the subject enters her choices). All eight models were estimated using Stata (version 12), and include a constant term and individual–subject random effects.

The table shows marginal effects (at variables’ means), standard errors, and level of significance for the main treatment variables. To save space, we leave the demographic and attitudinal variables, none of which were significant once a correction for multiple comparisons was made (Benjamini and Hochberg, 2005), out of the table, though we do keep these variables in the models themselves. Comparison of each model within a pair suggests that our main treatment variables are largely unaffected by whether these additional variables are included.

In Models 1 and 2, the marginal effects of the TG(0) and TG(0.5) dummies show that offers are significantly higher in the UG treatment than in either TG treatment, and further tests show that offers are also higher in TG(0.5) than in TG(0) ($p \approx 0.002$ and $p \approx 0.003$ in Models 1 and 2, respectively), consistent with Hypothesis 1. Model 3 further confirms the difference in offers between TG(0) and TG(0.5), though in Model 4, the difference misses significance ($p \approx 0.12$). Models 3 and 4 also show the positive correlation between messages and offers that was apparent from Figure 4, though calculations of the treatment-specific marginal effects show that this relationship is significant only for TG(0.5) ($p \approx 0.003$ and $p \approx 0.009$ in Models 3 and 4 respectively), and is insignificant and nearly zero for TG(0) ($p > 0.20$ in both models). The marginal effect of the round number is negative and

¹²Note that our TG(0) results are roughly comparable to those of Besancenot et al. (2013), who found truth telling in 8.8% of observations and “irrational lies” (i.e., in the “wrong” direction) in 2.7% of observations. They also found a negative relationship between offers and lies (see their Table 4), as the right panel of our Figure 4 indicates. Our model does not imply messages lower than offers (as Besancenot et al. remark about their irrational lies, “this strategy does not fit well in our theoretical framework”), and it is difficult to explain their non-negligible prevalence in the experiment. One possibility is that proposers believe they can “pleasantly surprise” the responder so as to get better treatment (Gneezy and List, 2006; Bellemare and Shearer, 2009). Another is that proposers simply believe that lower messages are more likely to be accepted (thinking perhaps that high messages will be perceived as blatant lies). However, regressions similar to those in Table 6 (available from the corresponding author) but distinguishing between exactly-truthful messages and under-statements find no significant evidence that under-statements are accepted by responders more often than truthful messages.

¹³Using linear instead of Tobit models does not qualitatively alter the conclusions (details available from the corresponding author).

Table 2: Proposer behaviour – Tobit marginal effects, with standard errors in parentheses

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Sample:	All proposers ($N = 425$)			TG(0) and TG(0.5) proposers ($N = 300$)				
Dependent variable:	Offer	Offer	Offer	Offer	Message	Message	Over–state	Over–state
TG(0) treatment	–0.220*** (0.041)	–0.225*** (0.043)						
TG(0.5) treatment	–0.096** (0.043)	–0.105** (0.044)	0.125*** (0.045)	0.089 (0.064)	–0.001 (0.023)	–0.010 (0.027)	–0.124*** (0.047)	–0.113* (0.060)
Round	–0.012*** (0.005)	–0.012** (0.005)	–0.018*** (0.005)	–0.018*** (0.006)	–0.007 (0.005)	–0.013** (0.006)	0.017*** (0.005)	0.013** (0.006)
Message			0.142** (0.071)	0.157** (0.075)				
Constant term?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional RHS variables?	No	Yes	No	Yes	No	Yes	No	Yes
$ \ln(L) $	43.76	59.42	2.87	18.85	161.32	184.51	39.74	19.38

Notes: RHS variables included but not displayed in Models 2, 4, 6 and 8: Living with family, Living with friends, Living alone, Older siblings, Younger siblings, Only child, Economics student, Business student, Number of economics classes completed, Age, Female, Decision time, 15 attitudinal variables (see Appendix C).

Offer, message, over–state as proportion (not percent) of cake size.

* (**, ***): Coefficient significantly different from zero at the 10% (5%, 1%) level.

significant in all four of these models, indicating the tendency of offers to decline over time on average, though this average masks differences across treatments: positive and insignificant time trend for UG, negative and significant for TG(0.5), negative and insignificant for TG(0).

Models 5 and 6 show that there is no treatment effect on messages, confirming the results of the non–parametric tests earlier in this section, and failing to support Hypothesis 5. Given the lack of treatment effect on messages and the significant treatment effects on offers seen in Models 1–4, we would expect over–statement to be larger in TG(0) than TG(0.5); this is confirmed by Models 7 and 8, again consistent with Hypothesis 6.

5.2 Responder behaviour

Table 3 shows some descriptive statistics for acceptances in four conditions: UG, TG(0.5) when the offer was seen by the responder, TG(0.5) when the offer was not seen, and TG(0). In the first two of these conditions, acceptance frequencies are broken down by intervals; in the last two, where offers were not observable, only overall acceptance frequencies are relevant. (Not shown, but easily calculated, is the overall acceptance frequency of 75% for TG(0.5).)

Responder behaviour in the UG, like that of proposers, is fairly typical. Equal splits are nearly always accepted, near–zero offers almost never, and intermediate offers with frequency in between. They respond similarly in TG(0.5) when the offer is visible, but with one important exception: they are significantly more likely to accept a low offer (less than one–fourth of the cake) than in the UG, though these are still accepted less often than higher offers. (As we will see shortly, this is because some of these low offers are accompanied by truthful messages.) They are also more likely to accept intermediate offers (25–45% of the cake), but this difference is not significant, and for high offers, there is no room for a treatment effect since they are already always accepted in the UG.

Consistent with Hypothesis 2, we see that the unconditional acceptance frequency in TG(0) is lower than that in either TG(0.5) or UG; however, only the difference between UG and TG(0) is significant, and only weakly so

Table 3: Responses to offers in UG, TG(0.5) and TG(0)

UG			TG(0.5) – offer observed			Signif. of diff.
Offers	Acceptances	Frequency (%)	Offers	Acceptances	Frequency (%)	
0–20%	0/5	0	0–20%	9/29	31	$p \approx 0.04$
25%–45%	66/92	72	25%–45%	35/45	78	n.s.
50%+	28/28	100	50%+	19/20	95	n.s.
All	94/125	75	All	63/94	67	n.s.
TG(0)			TG(0.5) – offer not observed			Signif. of diff.
Offers	Acceptances	Frequency (%)	Offers	Acceptances	Frequency (%)	
All	86/125	69	All	69/81	85	$p \approx 0.005$

See text for descriptions of significance tests. n.s. = not significant ($p > 0.20$).

(robust rank–order test, $p \approx 0.09$ for UG vs. TG(0), $p \approx 0.12$ for TG(0.5) vs. TG(0)). Similarly, a Kruskal–Wallis test cannot reject the null hypothesis that all three unconditional acceptance rates are equal ($p > 0.20$). On balance, then, the support for Hypothesis 2 is fairly weak.

Table 4 shows how responders respond to particular messages in the TG(0) and TG(0.5) treatments. Leaving aside the rare cases of very low (20% of the cake or less) or very high (75% or more) messages, responders are more likely to accept following unseen offers in TG(0.5) than following seen offers in that treatment, or following (unseen) offers in TG(0). Using Wilcoxon signed–ranks tests to compare TG(0.5) with unseen offers to TG(0.5)

Table 4: Responses to messages in TG(0) and TG(0.5)

Messages	TG(0)		TG(0.5) (offer seen)		TG(0.5) (offer unseen)	
	Acceptances	Frequency (%)	Acceptances	Frequency (%)	Acceptances	Frequency (%)
0–20%	3/7	43	2/2	100	2/3	67
25%–45%	28/43	65	27/35	77	21/23	91
50%–70%	48/67	72	33/55	60	44/52	85
75%–100%	7/8	88	1/2	50	2/3	67
All	86/125	69	63/94	67	69/81	85

with seen offers (matched samples), and robust rank–order tests to compare either of these with TG(0) (independent samples), messages between 25% and 45% of the cake are more likely to be accepted in TG(0.5) with an unseen offer than in TG(0) ($p \approx 0.026$), while messages between 50% and 70% of the cake in TG(0.5) are more likely to be accepted with an unseen than a seen offer ($p \approx 0.031$). Overall (pooling all messages), unseen offers in TG(0.5) are significantly more likely to be accepted than either seen offers in that treatment or offers in TG(0) (Wilcoxon signed–ranks test, $p \approx 0.016$ and robust rank–order test, $p \approx 0.010$ respectively).

Table 5 shows acceptance frequencies in those observations of TG(0.5) where the offer was seen by the responder, disaggregated according to whether the message was greater than the corresponding offer, or alternatively less than or equal to the offer; we continue to call a message a “lie” in the former case and “truthful” in the latter. The bottom row shows that in aggregate, offers are much more likely to be accepted when the accompanying message was truthful versus when it was a lie. This difference is significant (Wilcoxon signed–ranks test, $p \approx 0.008$), and consistent with Hypothesis 3. There appear to be significant differences even after conditioning on the offer, but the paucity of directly comparable observations (most truthful messages involve higher offers, while most lies involve

Table 5: Responders in TG(0.5) – offer observed

Message > Offer			Message ≤ Offer			Signif. of diff.
Offers	Acceptances	Frequency (%)	Offers	Acceptances	Frequency (%)	
0–20%	8/28	29	0–20%	1/1	100	n.s.
25%–45%	23/32	72	25%–45%	12/13	92	$p \approx 0.11$
50%+	0/0	–	50%+	19/20	95	–
All	31/60	52	All	32/34	94	$p \approx 0.008$

See text for descriptions of significance tests. n.s. = not significant ($p > 0.20$).

lower offers, so there is little overlap) means that the power is too low to reject a null of no difference, though the p -value for offers between 5 and 9 inclusive approaches statistical significance (Wilcoxon signed-ranks test, $p \approx 0.11$).

Not only is acceptance more likely in TG(0.5) after a truthful message than after a lie, but it is also more likely than in the UG: that is, when no message at all accompanied the offer. This can be seen by comparison between Tables 3 and 5. Moreover, the difference in aggregate acceptance frequencies is significant (robust rank-order test, $p \approx 0.003$).

We continue examining responder behaviour with several probit regressions, all with an Accept choice as the dependent variable. The models differ in which right-hand-side variables are included, which also determines the sub-sample used. Each model includes a constant term and the round number (though the latter turns out never to have a significant effect). Model 9 uses all of the responder data, and includes the offer and dummies for the TG(0) and TG(0.5) treatments (so that the baseline is UG). Model 10 uses the subset of the TG(0) and TG(0.5) data in which the offer was *unseen* (i.e., all observations in TG(0) and about half in TG(0.5)), and includes a TG(0) dummy along with the message. Model 11 uses the subset of the UG and TG(0.5) data in which the offer was *seen*, and includes the offer and dummy variables for a truthful message (taking a value of 1 if $m \leq x$) and a lie (equal to 1 if $m > x$); the baseline is thus the UG, where no message was sent. Model 12 uses the TG(0.5) data, and includes the message and a dummy for whether the offer was seen. Model 13 uses the subset of the TG(0.5) data in which the offer was seen, and includes the message and offer, along with the truthful-message dummy. These models additionally include all relevant two- and three-way interaction variables, though we leave out the demographic and attitudinal variables since we found no evidence that their inclusion in the proposer regressions (Table 2) had any significant effect on the results, and because the smaller sample sizes in some of these models may lead to concerns about over-fitting.

Three additional models look at the effects of previous-round lies, using two new indicator variables. The “lie caught in previous round” indicator takes a value of one if in the previous round, the responder was lied to by the matched proposer, but that proposer’s offer was observed (allowing the responder to react in the previous round). The “lie not caught in previous round” indicator takes a value of one if the responder was lied to and the offer was not observed (so the responder had no chance to react). The current message is also included on the right-hand side. Model 14 uses both TG(0) and TG(0.5) data where the current offer is unseen (and hence also includes an indicator for TG(0)), while Models 15 and 16 use the TG(0.5)–unseen and TG(0) data separately.

All of these models were estimated using Stata (version 12), and include individual-subject random effects. Table 6 reports the results.

Model 9 shows a positive and significant marginal effect of the offer on the probability of acceptance. Such a

Table 6: Responder behaviour – probit marginal effects (dependent variable = Accept), standard errors in parentheses

Model	9	10	11	12	13	14	15	16
Sample:	All responders	Both TG, offer unseen	UG and TG(0.5), offer seen	TG(0.5), all offers	TG(0.5), offer seen	Both TG, offer unseen	TG(0.5), offer unseen	TG(0) [offer unseen]
TG(0) treatment	0.162* (0.093)	-0.192*** (0.065)				-0.633* (0.372)		
TG(0.5) treatment	0.274*** (0.092)							
Offer (frac. of cake)	1.278*** (0.289)		2.052*** (0.742)		1.445 (1.261)			
Message (frac. of cake)		0.256 (0.219)		-0.269 (0.273)	0.982 (0.783)	0.853 (0.801)	0.000 (0.275)	0.445 (0.349)
Round	0.009 (0.022)	0.017 (0.023)	-0.011 (0.023)	0.030 (0.024)	0.010 (0.050)			
Offer seen				-0.218*** (0.064)				
Lie			0.063 (0.083)					
Truth			0.166** (0.076)		0.365*** (0.142)			
Lie caught in prev. round						-0.656 (0.447)	-0.123 (0.105)	
Lie not caught in prev. round						-0.758** (0.319)	-0.122 (0.157)	-0.264*** (0.097)
Constant term?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	425	206	219	175	94	167	67	100
$ \ln(L) $	218.40	105.47	86.44	88.59	35.31	86.22	25.54	60.28

* (**, ***): Coefficient significantly different from zero at the 10% (5%, 1%) level.

result may seem unsurprising given the strong correlation between offers and acceptance probabilities throughout the literature on experiments involving ultimatum games and related games; however, our finding is noteworthy since roughly half of all offers were not even observed by the responder. This model also shows positive and significant marginal effects for the TG(0) and TG(0.5) treatment dummies. This means that controlling for the offer, responders were on average more agreeable in these games than in the UG; the lower overall acceptance rate in TG(0) and roughly equal one in TG(0.5) seen earlier in the aggregate responder data are thus attributable mainly to lower offers by proposers, rather than to stricter responders.

Model 10 provides additional evidence consistent with Hypothesis 7; the negative and significant marginal effect of the TG(0) dummy implies that when an offer is not observed by the responder, it is more likely to be accepted if there was a chance that it could have been observed. Model 11 shows that controlling for the offer, the likelihood of acceptance following a lie in the TG(0.5) treatment is not significantly different from that in the UG, while acceptance is actually *more* likely after a truthful message. This suggests that responders are rewarding truthfulness rather than punishing lies, though this implication relies on the assumption that the UG (with no messages whatsoever, and hence no lies or truths) is an appropriate baseline.

In Model 12, the negative and significant marginal effect of the “offer seen” dummy indicates that responders are less likely to accept when they observe the offer than when they cannot observe it. The message itself has no significant effect on the likelihood of acceptance, despite the apparent positive correlation with the offer seen in

Figure 4 and Table 2. In Model 13, we further see that in cases where both offer and message are observed, not only the message, but also the offer has no significant effect on acceptance, except via the “truth” dummy. The positive and significant marginal effect of this last variable implies that responders like being told the truth, or alternatively dislike being lied to (consistent with Hypothesis 3), though the result for Model 11 suggests the former interpretation is more appropriate. By contrast, the lack of significance of either the message or the offer suggest that responders do not distinguish between big and little lies (or between big and little under-statements of the offer).

Models 14–16 provide further insight into how responders react to lies: “once bitten, twice shy”. Responders are less likely to accept based only on a message (i.e., after an unseen offer) if the previous message was a lie, relative to the baseline of having been told the truth. This effect is insignificant when the lie was caught in the previous round, but is significant in two of the three models when the lie had not been caught. (In the third model, the lack of significance is at least partly due to a small sample size.)

5.3 Payoffs

Table 7 shows the monetary payoffs of both player types in each cell of our experiment, along with p -values from Kruskal–Wallis tests of equal versus unequal payoffs across the three cells. The payoffs given acceptance illustrate

Table 7: Proposer and responder monetary payoffs (normalised to percents of cake size)

		UG	TG(0.5)	TG(0)	Significantly different?
Proposer payoff	(overall)	41.7	50.1	55.5	$p \approx 0.010$
	(given Accept)	55.4	66.4	80.7	$p \approx 0.008$
Responder payoff	(overall)	33.5	25.3	13.3	$p \approx 0.021$
	(given Accept)	44.6	33.6	19.3	$p \approx 0.008$

Note: Significance of differences based on Kruskal–Wallis 3-sample test.

how gains from bargaining are shared between the proposer and responder. These are most equitable in the UG, with proposers receiving just over half of the gains. Proposers’ shares rise to about two-thirds of gains in TG(0.5), and higher still (about 81 percent of gains) in TG(0). Despite the differences in acceptance rates across treatments, focusing on absolute payoffs yields the same relationships across treatments, with proposer payoffs lowest in UG and highest in TG(0), and the reverse ordering for responder payoffs. The Kruskal–Wallis tests confirm the significance of the differences across treatments.

6 Discussion

We have introduced a new game: the taxicab game. This game preserves the key features of the closely-related ultimatum game and yes/no game (Gehrig et al., 2007) – simple rules and an extremely asymmetric distribution of structural bargaining power – and incorporates a monetary incentive to lie, along with limited ability to detect a lie. We experimentally examine the taxicab game under two parameter specifications, along with the ultimatum game as a baseline, yielding three different environments: (i) TG(0), where the responder receives a message about the offer from the proposer, but not the offer itself (like the yes/no game but with a message); (ii) TG(0.5), where the message

is received, while the offer is observed with probability strictly between zero and one; and (iii) UG, where the offer is observed for sure, making it very similar to a TG(1) game (but with no message).

If all proposers and responders have standard “self-regarding” (*Homo economicus*) preferences, there should be no differences in behaviour across the three games: in each one, responders will accept any positive offer and perhaps even a zero offer, so proposers will offer nothing or nearly nothing, and messages will be irrelevant. However, many previous experimental results suggest that at least some agents are “other-regarding”, in that they are averse to inequity (advantageous and disadvantageous) and/or deception (lying and being lied to). When both self- and other-regarding agents are present in the population, our intuition (and a theoretical model presented in Appendix A) predict systematic differences in behaviour across the games, and across certain contingencies within a game.

To a large extent, these key differences are observed in our experiment. Our UG results look like ultimatum game results everywhere, with offers averaging about 40% of the cake and with low offers often rejected; this suggests that our subject pool and experimental procedures are fairly typical, thus providing a baseline from which to observe our other results. Our main results shed light on how individuals behave when they face monetary incentives to lie, and when they must interact with others who face such incentives. We find that lying by proposers is rampant but not universal: even in our TG(0) treatment, where lying cannot be detected until after all decisions are made, and even though our experimental procedures mean that subjects never face each other a second time, proposers choose not to over-state their offers about one-fifth of the time. Allowing the possibility for lying to be detected (our TG(0.5) treatment) doubles this frequency of “truth-telling”. More generally, increasing the likelihood that an offer is observed (from TG(0) to TG(0.5) to UG) results in higher offers and a decrease in the extent to which they are over-stated by messages.

Responders, for their part, behave in a way that combines an understanding of their own and their opponents’ monetary incentives with preferences over non-monetary aspects of outcomes. They frequently accept unseen offers, but do so more often when they know the offer was made with the recognition that it might be seen (the TG(0.5) treatment versus TG(0)); this makes sense given the beliefs responders in these situations are likely to have.¹⁴ When offers are seen, higher offers are of course more likely to be accepted, but the likelihood of accepting a given offer depends also on what message was sent. In particular, comparison of all observed offers – in both UG and TG(0.5) treatments – suggests that truthful messages are actually rewarded, in that they are more likely to be accepted than either lies or offers unaccompanied by a message. By contrast, we do not find that lies are punished in comparison to offers with no message. This pair of results is similar in spirit to results from other recent experimental studies showing that people are willing to sanction those who deceive them (and indeed those who deceive others, as mentioned in Section 2) relative to those who do not. However, to the extent that responders are rewarding good behaviour instead of punishing bad behaviour, our result is novel.¹⁵

Several potential extensions of our study immediately suggest themselves. Regarding theory, we view the model presented in Appendix A as a starting point, meant to justify our hypotheses as more than pure conjecture, rather than to argue for it as a literal description of reality. We expect that many other models incorporating both self- and other-regarding individuals can characterise the patterns of behaviour we saw in the experiment, and some may perform even better. One obvious possibility would be to give some responders a positive benefit from being told

¹⁴Specifically, if responders understand proposers’ incentive to make higher offers in TG(0.5) than TG(0) (to insure against rejection by other-regarding responders in case the offer is seen), they should have more optimistic beliefs about unseen offers in TG(0.5) than in TG(0).

¹⁵One admittedly speculative explanation for our apparent result of rewarding good behaviour is that responders in our experiment expect that they will be lied to, and are therefore pleasantly surprised upon receiving a truthful message, while subjects in similar positions in other experiments expected to be told the truth, and are disappointed upon finding out they were lied to. Such an explanation implies differences between our subject pool and those used in other experiments, which is possible but for which we have no evidence.

the truth, to complement or replace the disutility of being lied to. Other theoretical extensions could allow more heterogeneity in attitudes toward inequity and deception, or relax the assumption that individuals are affected in the same way from each of these – either by modifying the parameterisation of our model or within a different model.

Other extensions would incorporate both theory and experiments. One could look at different ultimatum games with p -observability and with/without messages. How does the behaviour of proposers and responders change as p varies between 0 and 1? Are there any noteworthy differences between TG(1) and the standard ultimatum game? How robust is our apparent result of a truth-telling premium; does it exist for all p , or only for p sufficiently low that proposers could reasonably expect to “get away” with lying, so that truth telling could be attributed to personal ethics rather than a fear of getting caught? Alternatively, one could keep the same value of p , but vary the disagreement outcome. As mentioned earlier (see footnote 1), making the responder’s disagreement payoff positive instead of zero changes the game substantially for self-regarding responders (and arguably would be in the spirit of the taxicab example described in our introduction). A third extension would investigate the taxicab game with proposer competition. If responders can choose among proposers based on the messages they receive – and with some known probability, the true offers – how will they choose, and how will this choice impact on the offers and messages proposers choose? Still other extensions could examine how behaviour in the taxicab game is affected by institutions such as repetition with fixed pairings, reputation mechanisms, second- or third-party punishment, or legal remedies.

There are likely many other interesting extensions. We expect the taxicab game will become useful as a milieu for understanding how individuals trade off between honesty and monetary gain when providing information to others, and how they react to potentially fraudulent information in order to deter dishonest information and encourage truth-telling. Such situations are ubiquitous in economic (and other) settings.

References

- Abeler, J., A. Becker and A. Falk (2014), “Representative evidence on lying costs”, *Journal of Public Economics* 113, pp. 96–104.
- Aimone, J.A., S. Ball and B. King-Casas (2013), “Social and economic risk: results from an individual level betrayal aversion measure”, working paper, Virginia Tech.
- Aimone, J.A. and D. Houser (2012), “What you don’t know won’t hurt you: a laboratory analysis of betrayal aversion”, *Experimental Economics* 15, pp. 571–588.
- Babcock, L., G. Loewenstein, S. Issacharoff and C. Camerer (1995), “Biased judgments of fairness in bargaining”, *American Economic Review* 85, pp. 1337–1343.
- Banks, J.S. (1990), “A model of electoral competition with incomplete information”, *Journal of Economic Theory* 50, pp. 309–325.
- Bellemare, C. and B. Shearer (2009), “Gift giving and worker productivity: evidence from a firm-level experiment”, *Games and Economic Behavior* 67, pp. 233–244.
- Besancenot, D., D. Dubart and R. Vranceanu (2013), “The value of lies in an ultimatum game with imperfect information”, *Journal of Economic Behavior & Organization* 93, pp. 239–247.

- Benjamini, Y. and Y. Hochberg (1995), “Controlling the false discovery rate: a practical and powerful approach to multiple testing”, *Journal of the Royal Statistical Society, Series B (Methodological)* 57, pp. 289–300.
- Binmore, K., C. Proulx, L. Samuelson and J. Swierzbinski (1998), “Hard bargains and lost opportunities”, *Economic Journal* 108, pp. 1279–1298.
- Blanco, M., D. Engelmann and H.-T. Normann (2011), “A within–subject analysis of other regarding preferences”, *Games and Economic Behavior* 72, pp. 321–338.
- Bohnet, I. and R. Zeckhauser (2004), “Trust, risk and betrayal”, *Journal of Economic Behavior & Organization* 55, pp. 467–484.
- Bolton, G.E. and A. Ockenfels (2000), “ERC: a theory of equity, reciprocity, and competition”, *American Economic Review* 90, pp. 166–193.
- Brandts, J. and G. Charness (2003), “Truth or consequences: an experiment”, *Management Science* 49, pp. 116–130.
- Cadsby, B., N. Du, F. Song and L. Yao (2013), “Promise–keeping, social identity, and social distance: an experimental investigation”, working paper.
- Camerer, C.F. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press, Princeton, NJ.
- Charness, G. and M. Dufwenberg (2006), “Promises and partnership”, *Econometrica* 74, pp. 1579–1601.
- Charness, G. and M. Dufwenberg (2010), “Bare promises: an experiment”, *Economics Letters* 107, pp. 281–283.
- Chen, Y., N. Kartik and J. Sobel (2008), “Selecting cheap–talk equilibria”, *Econometrica* 76, pp. 117–136.
- Cho, I.-K. (1987), “A refinement of sequential equilibrium”, *Econometrica* 55, pp. 1367–1389.
- Corazzini, L., S. Kube, M.A. Maréchal and A. Nicolò (2013), “Elections and deceptions: an experimental study on the behavioral effects of democracy”, forthcoming, *American Journal of Political Science*.
- Croson, R.T. (1996), “Information in ultimatum games: an experimental study”, *Journal of Economic Behavior & Organization* 30, pp. 197–212.
- Croson, R., T. Boles and J.K. Murnighan (2003), “Cheap talk in bargaining experiments: lying and threats in ultimatum games”, *Journal of Economic Behavior & Organization* 51, pp. 143–159.
- de Laplace, P.S. (1814), *Théorie Analytique des Probabilités*, Paris.
- Dufwenberg, M. and U. Gneezy (2000), “Measuring beliefs in an experimental lost wallet game”, *Games and Economic Behavior* 30, pp. 163–182.
- Ellingsen, T. and M. Johannesson (2004). “Promises, threats and fairness”, *Economic Journal* 114, pp. 397–420.
- Ellingsen, T., M. Johannesson, S. Tjøtta and G. Torsvik (2010), “Testing guilt aversion”, *Games and Economic Behavior* 68, pp. 95–107.

- Fehr, E., and K.M. Schmidt (1999), “A theory of fairness, competition and cooperation”, *Quarterly Journal of Economics* 114, pp. 817–868.
- Feltovich, N. (2005), “Critical values for the robust rank–order test”, *Communications in Statistics—Simulation and Computation* 34, pp. 525–547.
- Fischbacher, U. (2007), “z–Tree: Zurich toolbox for ready–made economic experiments”, *Experimental Economics* 10, pp. 171–178.
- Fischbacher, U. and Föllmi-Heusi (2013), “Lies in disguise – an experimental study on cheating”, *Journal of the European Economic Association* 11, pp. 525–547.
- Gehrig, T., W. Güth, V. Levati, R. Levinsky, A. Ockenfels, T. Uske and T. Weiland (2007), “Buying a pig in a poke: an experimental study of unconditional veto power”, *Journal of Economic Psychology* 28, pp. 692–703.
- Gneezy, U. (2005). “Deception: the role of consequences”, *American Economic Review* 95(1), pp. 384–394.
- Gneezy, U. and J.A. List (2006), “Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments”, *Econometrica* 74, pp. 1365–1384.
- Greiner, B. (2004), “An online recruitment system for economic experiments”, in Kremer, K. and V. Macho eds., *Forschung und wissenschaftliches Rechnen. GWDG Bericht 63*. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung, pp. 79–93.
- Güth, W., and O. Kirchkamp (2012), “Will you accept without knowing what? The Yes–No game in the newspaper and in the lab”, *Experimental Economics* 15, pp. 656–666.
- Hurkens, S. and N. Kartik (2009), “Would I lie to you? On social preferences and lying aversion”, *Experimental Economics* 12, pp. 180–192.
- Kagel, J., C. Kim and D. Moser (1996), “Fairness in ultimatum games with asymmetric information and asymmetric payoffs”, *Games and Economic Behavior* 13, pp. 100–110.
- Kartik, N. (2009), “Strategic communication with lying costs”, *Review of Economic Studies* 76, pp. 1359–1395.
- Kriss, P.H., R. Nagel and R.A. Weber (2013), “Implicit vs. explicit deception in ultimatum games with incomplete information”, forthcoming, *Journal of Economic Behavior & Organization*.
- Lang, K. and R.W. Rosenthal (1991), “The contractors’ game”, *RAND Journal of Economics* 22, pp. 329–338.
- López-Pérez, R. (2012), “The power of words: a model of honesty and fairness”, *Journal of Economic Psychology* 33, pp. 642–658.
- López-Pérez, R. and E. Spiegelman (2013), “Why do people tell the truth? Experimental evidence for pure lie aversion”, forthcoming, *Experimental Economics*.
- Lundquist, T., T. Ellingsen, E. Gribbe and M. Johannesson (2009), “The aversion to lying”, *Journal of Economic Behavior & Organization* 70, pp. 81–92.

- McKelvey, R.D. and T.R. Palfrey (1995), “Quantal response equilibria for normal form games”, *Games and Economic Behavior* 10, pp. 6–38.
- Miettinen, T. (2013). “Promises and conventions: an approach to pre–play agreements”, *Games and Economic Behavior*
- Mitzkewitz, M. and R. Nagel (1993), “Experimental results on ultimatum games with incomplete information”, *International Journal of Game Theory* 22, pp. 171–198.
- Morris, S. (1995), “The common prior assumption in economic theory”, *Economics and Philosophy* 11, pp. 227–253.
- Ohtsubo, Y., F. Masuda, E. Watanabe and A. Masuchi (2010), “Dishonesty invites costly third–party punishment”, *Evolution and Human Behavior* 31, pp. 259–264.
- Peeters, R., M. Vorsatz and M. Walzl (2008), “Rewards in an experimental sender–receiver game”, *Economics Letters* 101, pp. 148–150.
- Roth, A.E. (1995), “Bargaining experiments”, in *Handbook of Experimental Economics*, J.H. Kagel and A.E. Roth, eds., Princeton University Press, Princeton, NJ, pp. 253–348.
- Roth, A. and J.K. Murnighan, J.K. (1982). “The role of information in bargaining: an experimental study”, *Econometrica* 50, pp. 1123–1142.
- Roth, A. and J.K. Murnighan, J.K. (1983), “Information and aspirations in two–person bargaining”, in Tietz, R. ed., *Aspiration Levels in Bargaining and Economic Decision Making*. New York: Springer, pp. 91–103.
- Sánchez-Pagés, S. and M. Vorsatz (2007), “An experimental study of truth–telling in a sender–receiver game”, *Games and Economic Behavior* 61, pp. 86–112.
- Sánchez-Pagés, S. and M. Vorsatz (2009), “Enjoy the silence: an experiment on truth–telling”, *Experimental Economics* 12, pp. 220–241.
- Siegel, S. and N.J. Castellan, Jr. (1988), *Nonparametric Statistics for the Behavioral Sciences*, MTGraw–Hill, New York.

A Theoretical analysis

Here, we develop and analyse a model of behaviour in UG, TG(0.5) and TG(0). Such an analysis is trivial under the common assumption of self–regarding individuals, but we consider the possibility that at least some are other–regarding, along the lines of Fehr and Schmidt’s (1999) model of *inequity aversion*, according to which players dislike both advantageous inequity (getting a higher payoff than others) and disadvantageous inequity (getting a lower payoff than others). Under the two–player version of their model, utility is given by

$$U_i(x) = x_i - \alpha_i \cdot \text{Max}\{x_j - x_i, 0\} - \beta_i \cdot \text{Max}\{x_i - x_j, 0\},$$

where x_i is the own payoff and x_j is the opponent payoff, $\alpha_i \geq \beta_i \geq 0$, and $\beta_i < 1$. The values of α_i and β_i represent the player’s aversion to disadvantageous and advantageous inequity, respectively. If we restrict consideration to offers of 50% or less of the cake, the utility functions simplify to

$$U_p(x) = 20 - x - \beta_p(20 - 2x); \quad (1)$$

$$U_r(x) = x - \alpha_r(20 - 2x); \quad (2)$$

for proposers and responders respectively. These are the utility functions we use for the ultimatum game.¹⁶

We also consider players who dislike deception. For proposers in either version of the taxicab game, this is an aversion to lying, while for responders it is an aversion to being lied to. When $m > x$ (the message over-states the offer) and the responder accepts, the proposer receives a disutility of $\gamma_p(m - x)$ and the responder receives a disutility of $\delta_r(m - x)$, on top of any money payoff and inequity aversion. When $x \geq m$, disutility from deception is zero, as in games like the UG where no messages are sent. Disutility from deception is also zero when the responder rejects, since in that case no harm was done. Adding deception aversion to (1) and (2) gives us our general utility functions:

$$U_p(x, m) = \begin{cases} 20 - x - \beta_p(20 - 2x) - \gamma_p \cdot \text{Max}\{0, m - x\} & x \text{ accepted;} \\ 0 & x \text{ rejected;} \end{cases} \quad (3)$$

$$U_r(x, m) = \begin{cases} x - \alpha_r(20 - 2x) - \delta_r \cdot \text{Max}\{0, m - x\} & x \text{ accepted;} \\ 0 & x \text{ rejected;} \end{cases} \quad (4)$$

for $x, m \leq 10$. We call a player “other-regarding” if at least one of $\alpha_r, \beta_p, \gamma_p$ or δ_r is strictly positive, and “self-regarding” if all are zero.

A.1 A simple model of behaviour in the UG and TG games

We now make several assumptions in an attempt to simplify the general model while still yielding non-trivial predictions for the experiment. We assume that the population comprises both self-regarding and other-regarding people, with $\phi_s \in (0, 1)$ the fraction of self-regarding proposers and $\rho_s \in (0, 1)$ the fraction of self-regarding responders (with complements $\phi_o = 1 - \phi_s$ and $\rho_o = 1 - \rho_s$). All self-regarding people are obviously identical, but we also assume that all of the other-regarding people are identical, with $\alpha = \beta = \gamma = \delta = 0.5$. Such an assumption is strong, but perhaps justifiable by an argument that individuals either value “fair play” or they do not, and those who value fair play ought to dislike both greed and lying, while those who do not are indifferent toward both.¹⁷ The values of ϕ_s and ρ_s are assumed to be common knowledge. The common value of one-half for all of the parameters is also an obvious simplification, though in the case of inequity aversion, it is close to the median values reported by Blanco, Engelmann and Normann (2011), and some of our results are robust to relaxing this assumption.¹⁸ Let P_s and P_o stand for self- and other-regarding proposers respectively, and similarly R_s and R_o for responders. Define

¹⁶Restricting consideration to offers of half the cake or less is not completely without loss of generality, since subjects in the experiment can make larger offers. However, in ultimatum game experiments, proposers seldom offer more than half the cake, and we conjecture that such offers will be no more common, indeed perhaps less common, in the taxicab game. We also restrict consideration to messages of 50% or less of the cake; since offers are unlikely to be outside this range, messages are likely to be viewed as obviously incredible if they indicate such a high offer.

¹⁷Indeed, this simplification may not be unrealistic. For example, Hurkens and Kartik (2009) argue that the data from Gneezy’s (2005) lying experiment are explained well by a model with only two types: those who lie whenever it benefits them, and those who never lie.

¹⁸For example, other-regarding proposers’ behaviour in each game would be unaffected by increasing the values of any of the parameters.

$\mu_s(m)$ to be the responder's belief that a given proposer is self-regarding based on observing m in either version of the taxicab game, and let $\mu_o(m) = 1 - \mu_s(m)$ be the corresponding belief that the proposer is other-regarding.

We next make three tie-breaking assumptions. First, if an other-regarding proposer is indifferent between multiple messages, she will choose the one that minimises over-statement ($\text{Max}\{0, m-x\}$), and if multiple messages obtain this minimum (as happens when $x > 0$), she will choose the most truthful one (i.e., the closest to x). Second, if a proposer is indifferent between multiple offers, all of which are less than or equal to half of the cake, she will send the highest offer. Third, if a responder is indifferent between accepting or rejecting, he will accept.

Under these assumptions, the basic taxicab game $\text{TG}(0)$ becomes a *signalling game with multiple sender and receiver types*, in contrast to standard signalling games that have multiple sender types but just one receiver type. The modified taxicab game $\text{TG}(0.5)$ is a *semi-signalling game* in which nature reveals information about the sender's type to the receiver with probability one-half but does not reveal the receiver's type to the sender.

A.2 Equilibria

The solutions of the games will depend on the population proportions of self-regarding proposers and responders (which we simply call "model parameters", as we have fixed the values of all other parameters).

Proposition 1 *The UG has a unique perfect Bayesian equilibrium (PBE). Other-regarding responders will accept offers of at least 5, and self-regarding responders will accept all offers. Other-regarding proposers will offer 10, and self-regarding proposers will offer either 0 (if $\rho_o < 0.25$) or 5 (otherwise).*

Proof: The solution is by backward induction. Self-regarding responders will accept any offer. The utility function of an other-regarding responder is given by $U_{r_o}(x) = x - 0.5(20 - 2x) = 2x - 10$ if he accepts the proposer's offer, and 0 if he rejects. So, he will accept any offer of at least 5. Thus, proposers know that any offer $x \geq 5$ will be accepted for sure, and any lower offer will be accepted with probability ρ_s (the fraction of self-regarding responders in the population).

A self-regarding proposer will offer either 5 or 0, depending on ρ_s . If $\rho_s > 0.75$, she will offer 0; otherwise she will offer 5 (by our second tie-breaking rule when $\rho_s = 0.75$). An other-regarding proposer has the utility function $U_{p_o}(x) = 20 - x - 0.5(20 - 2x) = 10$, so prefers to offer 10 (by our second tie-breaking rule) as long as it is accepted (which it is).

Unlike the UG, $\text{TG}(0)$ has no equilibrium in which self-regarding proposers make positive offers.

Proposition 2 *In $\text{TG}(0)$, no separating PBE exists. There exists at least one **pooling PBE**, with all proposers sending the same message, other-regarding proposers offering 10, and self-regarding proposers offering 0. When $\phi_o \geq 0.5$, there is a pooling PBE where all offers are accepted, and any message between 0 and 10 is consistent with equilibrium. When $\phi_o < 0.6$, there is a pooling PBE where only messages of 10 are sent, and only self-regarding responders accept offers.*

Note that for $\phi_o \in [0.5, 0.6)$, the two PBE co-exist.¹⁹

Proof:

[Part 1: no separating equilibrium]

¹⁹We note here that the yes/no game has equilibria very similar to $\text{TG}(0)$. There is no separating equilibrium, and there is always a pooling equilibrium where self-regarding proposers offer 0, other-regarding proposers offer 10, self-regarding responders always accept, and other-regarding responders accept if $\phi_o \geq 0.5$.

Suppose by contradiction there does exist a separating equilibrium, with self-regarding proposers choosing (x_s, m_s) , other-regarding proposers choosing (x_o, m_o) , and $m_o \neq m_s$. Then responder beliefs after m_s or m_o are given by $\mu_s(m_s) = 1$ and $\mu_s(m_o) = 0$. Also, self-regarding responders will always accept, and self-regarding proposers will offer 0, so that other-regarding responders will reject after receiving m_s . If other-regarding responders accept after observing m_o then the payoff for other-regarding proposers is

$$(20 - x_o) - 0.5(20 - 2x_o) - 0.5 \cdot \text{Max}(0, m_o - x_o) = \begin{cases} 10 - 0.5(m_o - x_o) & x_o \leq m_o \\ 10 & x_o > m_o. \end{cases}$$

On the other hand, if they reject after observing m_o the payoff for other-regarding proposers is

$$[(20 - x_o) - 0.5(20 - 2x_o) - 0.5 \cdot \text{Max}(0, m_o - x_o)]\rho_s = \begin{cases} [10 - 0.5(m_o - x_o)]\rho_s & x_o \leq m_o \\ 10\rho_s & x_o > m_o. \end{cases}$$

By assumption, $m_o \leq 10$. So, from the payoff functions specified above, other-regarding proposers will choose $x_o = 10$ in this potential separating equilibrium regardless of the strategy of other-regarding responders. Hence, in this equilibrium, other-regarding responders must accept after observing m_o . But then, self-regarding proposers will deviate and choose $(x_s, m_s) = (0, m_o)$ breaking the equilibrium. So, such an equilibrium cannot exist.

[Part 2: pooling equilibrium]

Consider a pooling equilibrium, where both types of proposer send message m^* and other-regarding proposers offer x^* . Then responder beliefs after seeing m^* must be $\mu_s = \phi_o$; other messages are sent with probability zero, so we can assume $\mu_s(m) = 1$ for $m \neq m^*$. Self-regarding responders will accept after any message. Self-regarding proposers will offer 0, irrespective of what message they send.

Given proposers' behaviour, an other-regarding responder receiving a message m^* believes he is getting

$$\begin{aligned} & \phi_o[x^* - 0.5(20 - 2x^*) - 0.5 \cdot \text{Max}(0, m^* - x^*)] + (1 - \phi_o)[0 - 0.5(20) - 0.5(m^*)] \\ &= \begin{cases} 2.5\phi_o x^* - 10 - 0.5m^* & x^* \leq m^* \\ 2\phi_o x^* - 10 - 0.5(1 - \phi_o)m^* & x^* > m^* \end{cases} \end{aligned}$$

in expected utility if he accepts, and 0 if he rejects. He will accept if either (a) $x^* \leq m^*$ and $x^* \geq \frac{m^*+20}{5\phi_o}$ or (b) $x^* > m^*$ and $x^* \geq \frac{(1-\phi_o)m^*+20}{4\phi_o}$. That is, he accepts if $x^* \geq \text{Max} \left\{ \frac{m^*+20}{5\phi_o}, \frac{(1-\phi_o)m^*+20}{4\phi_o} \right\}$. After any other message, he would believe he is facing a self-regarding proposer who offered zero, and would reject (since the utility from accepting is $-0.5m - 10 < 0$).

First, we look for pooling PBE in which other-regarding responders accept after observing m^* . In this case, the other-regarding proposer's utility is equal to $10 - 0.5(m^* - x_o)$ if $x_o \leq m^*$, or 10 if $x_o > m^*$. Given that $m^* \leq 10$, an other-regarding proposer will choose $x_o = 10$, by our second tie-breaking rule. Then, the condition of acceptance by other-regarding responders will hold if $10 \geq \text{Max} \left\{ \frac{m^*+20}{5\phi_o}, \frac{(1-\phi_o)m^*+20}{4\phi_o} \right\}$, which holds if $\phi_o \geq \frac{20+m^*}{40+m^*}$. As Table 8 shows, as long as $\phi_o \geq 0.5$, there exists a combination of m^* and ϕ_o that supports a pooling PBE in which other-regarding responders accept after observing m^* . For $\phi_o < 0.5$, it must be that

Table 8: Minimum proportion of other-regarding responders needed to support a given message in a PBE of TG(0)

m^* :	0	1	2	3	4	5	6	7	8	9	10
ϕ_o :	0.5	0.512	0.524	0.535	0.545	0.556	0.565	0.574	0.583	0.592	0.6

$10 < \text{Max} \left\{ \frac{m^*+20}{5\phi_o}, \frac{(1-\phi_o)m^*+20}{4\phi_o} \right\}$, so no PBE of this form exists.

Finally, we look for pooling PBE in which other-regarding responders reject irrespective of the message. In that case, self-regarding proposers will offer 0 and other-regarding proposers will offer 10 (by the second tie-breaking rule) and send a message of 10 (by the first tie-breaking rule), while self-regarding responders will still accept irrespective of the message. As long as $\phi_o < 0.6$, other-regarding responders do prefer to reject after a message of 10 (by the same reasoning that yielded Table 8), completing the equilibrium.

In solving TG(0.5), we introduce some additional terminology. An equilibrium in which both proposer types send the same message and make the same offer is a *pooling equilibrium*, one where both proposer types send the same message but make different offers is a *semi-pooling equilibrium*, and one where both proposer types send different messages and make different offers is a *separating equilibrium*.

Proposition 3 *In TG(0.5), no pooling PBE exist. There always exists a semi-pooling PBE with other-regarding proposers offering 10, but for a small subset of parameter values ($\rho_o \in [0.4, 0.5)$ and $\phi_o < 0.5$) there is none where everyone chooses pure strategies. For $\rho_o \geq 0.5$, both semi-pooling and separating PBE exist, with messages between 0 and 10, self-regarding proposers offering 5 or 6 (depending on model parameters and the message(s)), and all offers being accepted. For $\rho_o < 0.5$ and $\phi_o \geq 0.5$, a semi-pooling PBE exists, with all proposers sending the same message, self-regarding proposers offering 0, and all offers being accepted except 0 offers by other-regarding responders who see both the message and the offer. For $\rho_o < 0.4$ and $\phi_o \leq 0.6$, a semi-pooling PBE exists with all proposers sending a message of 10, self-regarding proposers offering 0, self-regarding responders always accepting, and other-regarding responders accepting only when they see the offer of 10.*

Proof:

[Part 1: no pooling equilibrium]

Suppose that a pooling equilibrium does exist; that is, both types of proposer send message m^* and offer x^* . Then self-regarding responders will always accept, and other-regarding responders will accept after any combination (x, m) with $x \geq \text{Max} \left\{ 5, 4 + \frac{m}{5} \right\}$, and will accept (\emptyset, m^*) if $x^* \geq \text{Max} \left\{ 5, 4 + \frac{m^*}{5} \right\}$.

In a potential pooling equilibrium, if other-regarding responders accept after observing m^* then the payoff for other-regarding proposers is

$$(20 - x^*) - 0.5(20 - 2x^*) - 0.5 \cdot \text{Max}(0, m^* - x^*) = \begin{cases} 10 - 0.5(m^* - x^*) & x^* \leq m^* \\ 10 & x^* > m^*. \end{cases}$$

On the other hand, if they reject after observing m^* the payoff for other-regarding proposers is

$$[(20 - x^*) - 0.5(20 - 2x^*) - 0.5 \cdot \text{Max}(0, m^* - x^*)]\rho_s = \begin{cases} (10 - 0.5(m^* - x^*))\rho_s & x^* \leq m^* \\ 10\rho_s & x^* > m^*. \end{cases}$$

As long as $m^* \leq 10$; this is maximised when $x^* \geq m^*$, so by the second tie-breaking rule, we have $x^* = 10$. So, in any pooling equilibrium, it must be that $x^* = 10$. However, a self-regarding proposer would earn 10 by following this strategy, and would earn $20(1 - \rho_o/2) > 10$ by offering zero instead, breaking the potential equilibrium.

[Part 2: separating equilibrium]

In a separating PBE, self-regarding proposers choose (x_s, m_s) and other-regarding proposers choose (x_o, m_o) , with $m_s \neq m_o$. Self-regarding responders always accept, and other-regarding responders accept after seeing (x, m) if $x \geq \text{Max} \left\{ 5, 4 + \frac{m}{5} \right\}$, after (\emptyset, m_o) if $x_o \geq \text{Max} \left\{ 5, 4 + \frac{m_o}{5} \right\}$, and after (\emptyset, m_s) if $x_s \geq \text{Max} \left\{ 5, 4 + \frac{m_s}{5} \right\}$.

First, there is no separating PBE where other-regarding responders reject after (\emptyset, m_o) . Since self-regarding proposers are choosing a different message, this could only happen if $x_o < \text{Max} \{5, 4 + \frac{m_o}{5}\}$, but in this case, then other-regarding proposers would offer 10 (by the second tie-breaking rule) instead, making other-regarding responders prefer to accept.

Second, there is no separating PBE where other-regarding responders reject after (\emptyset, m_s) but accept after (\emptyset, m_o) . In this case, the best that a self-regarding proposer could do in such a PBE is by offering zero (maximising payoff in case she is matched with a self-regarding responder, since she earns nothing from being matched with an other-regarding responder). This would give her $20\rho_s = 20(1 - \rho_o)$. But deviating to message m_o , while still offering zero, would lead to acceptance except when matched with an other-regarding responder who saw the offer (since outside this case, he would infer that it was an other-regarding proposer), yielding payoff $20 \left(1 - \frac{1}{2}\rho_o\right)$, which is strictly larger, breaking the equilibrium.

Third, consider a separating equilibrium where other-regarding responders accept after both (\emptyset, m_o) and (\emptyset, m_s) . Then other-regarding proposers will offer 10. Self-regarding proposers will offer the lowest amount that gets accepted. From $x_s \geq \text{Max} \{5, 4 + \frac{m_s}{5}\}$ above, it follows that $x_s = 5$ if $m_s \leq 5$, and $x_s = 6$ if $m_s \geq 6$. It remains to confirm that this is actually a best response for them; that is, it would not pay to deviate by offering 0. Offering 0 earns $20 \left(1 - \frac{1}{2}\rho_o\right)$, while offering 5 (6) earns 15 (14) for sure; the latter is higher if $\rho_o \geq 0.5$ ($\rho_o \geq 0.6$). Thus, it is an equilibrium for other-regarding proposers to choose $(10, m_s)$ and self-regarding proposers to choose $(5, m_o)$ if $m_s \neq m_o$, $m_s \leq 5$ and $\rho_o \geq 0.5$, and it is an equilibrium for other-regarding proposers to choose $(10, m_s)$ and self-regarding proposers to choose $(6, m_o)$ if $m_s \neq m_o$, $m_s \geq 6$ and $\rho_o \geq 0.6$. If $\rho_o < 0.5$, there is no equilibrium of this type.

When a separating equilibrium exists, it is supported by beliefs that the proposer is self-regarding and offered zero after any out-of-equilibrium (offer, message) pair (\emptyset, m) .

[Part 3: semi-pooling equilibrium]

In a semi-pooling equilibrium, both proposers choose message m^* ; self-regarding proposers offer x_s and other-regarding proposers offer x_o , with $x_o \neq x_s$. Self-regarding responders always accept, and other-regarding responders accept after seeing (x, m) if $x \geq \text{Max} \{5, 4 + \frac{m}{5}\}$. Then, other-regarding proposers will choose $x_o = 10$. Given (offer, message) pair (\emptyset, m^*) , an other-regarding responder will get a payoff of 0 from rejecting, and accepting will yield $(1 - \phi_s)10 + \phi_s(2x_s - 10)$ if $x_s \geq m^*$, or $(1 - \phi_s)10 + \phi_s(2.5x_s - 10 - 0.5m^*)$ if $x_s < m^*$. So, he will accept if $x_s \geq \text{Max} \left\{10 - \frac{5}{\phi_s}, 8 - \frac{4}{\phi_s} + \frac{m^*}{5}\right\}$. Note that this is a strictly weaker condition than $x_s \geq \text{Max} \{5, 4 + \frac{m}{5}\}$, since $\phi_s < 1$.

Given that other-regarding responders accept following (\emptyset, m^*) , a self-regarding proposer has two possible optimal offers. Offering 0 earns an expected payoff of $20 - 10\rho_o$, and offering the minimum x that satisfies $x \geq \text{Max} \left\{5, 4 + \frac{m^*}{5}\right\}$ (either 5 or 6, depending on m^*) earns $20 - x$. For $m^* \leq 5$, this minimum x is 5, which is payoff-maximising if $\rho_o \geq 0.5$. For $m^* \geq 6$, this minimum x is 6, which is payoff-maximising if $\rho_o \geq 0.6$. Note that other-regarding responders will accept after receiving (\emptyset, m^*) if $x_s \geq 5, m^* \leq 5$ or if $x_s \geq 6, m^* \geq 6$. In other cases, self-regarding proposers will offer 0, which might still be consistent with equilibrium, as long as other-regarding responders accept, which happens if $0 \geq \text{Max} \left\{10 - \frac{5}{\phi_s}, 8 - \frac{4}{\phi_s} + \frac{m^*}{5}\right\}$, that is, $\phi_s \leq \frac{20}{40 + m^*}$.

Alternatively, suppose $x_s < \text{Max} \left\{10 - \frac{5}{\phi_s}, 8 - \frac{4}{\phi_s} + \frac{m^*}{5}\right\}$, so that other-regarding responders will reject following (\emptyset, m^*) . Then, it must be that they reject following all (\emptyset, m) (otherwise, proposers would choose a message that led to acceptance). So, other-regarding proposers will choose $(10, 10)$, so that it must be that $m^* = 10$. If self-regarding proposers make an offer acceptable to other-regarding responders, this will break the equilibrium, so

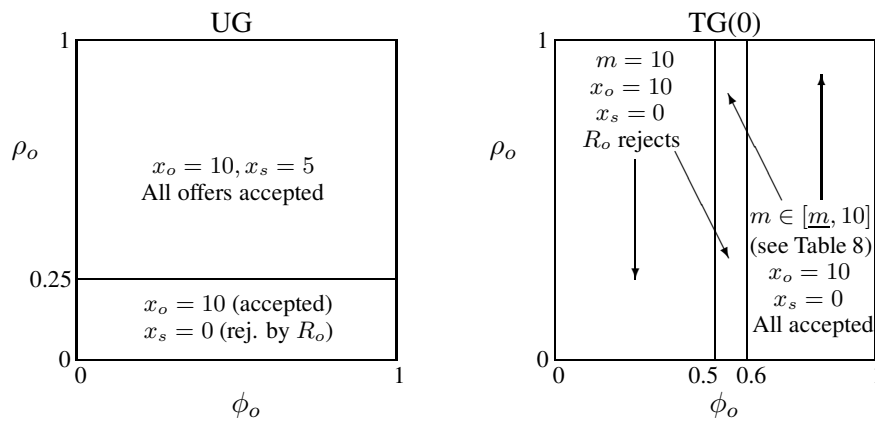
it must be that they offer 0, along with sending message 10. This earns $20(1 - \rho_o)$, while deviating to an offer of 5 (the lowest possible acceptable offer, when combined with a message of 5 or less) earns $15(1 - \rho_o/2)$. The offer of 0 is payoff-maximising if $\rho_o \leq 0.4$, and consistent with rejection by other-regarding responders if $\phi_s > 0.4$.

For $\phi_o < 0.5$ and $\rho_o \in [0.4, 0.5)$, there is no pure-strategy equilibrium, but there are semi-pooling PBE where some players mix. So, for example, it is easily verified that there is an equilibrium where both types of proposer send message $m^* \leq 5$, other-regarding proposers offer 10, self-regarding proposers mix between offering 0 and 5, self-regarding responders always accept, and other-regarding responders accept following $(5, m^*)$ or $(10, m^*)$, reject following $(0, m^*)$, and mix between accepting and rejecting following (\emptyset, m^*) .

When a semi-pooling equilibrium exists, it is supported by beliefs that the proposer is self-regarding and offered zero after any out-of-equilibrium pair (\emptyset, m) .

Figures 5 and 6 show how the PBE of the three games depend on the proportions of other-regarding players in the proposer and responder populations.

Figure 5: Equilibrium behaviour in UG and TG(0), conditional on ϕ_o and ρ_o



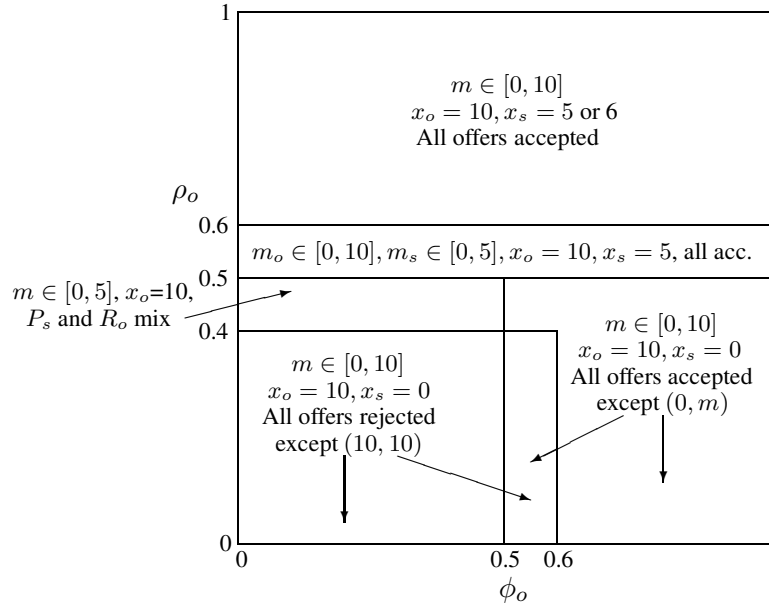
A.3 Implications for the experiment

From Propositions 1, 2 and 3, we see that equilibrium behaviour depends on the values of ϕ_o and ρ_o – that is, the proportions of other-regarding proposers and responders – and even for specific values, there can be multiple perfect Bayesian equilibria. We therefore need to make two additional assumptions in order to arrive at the predictions we will test in the experiment.

Our first assumption is that, since subjects in the experiment were randomly assigned to their roles (proposer or responder), role should be uncorrelated with behavioural type, and so the proportion of other-regarding proposers should be the same as that of other-regarding responders: $\phi_o = \rho_o$.²⁰ Then the set of equilibria can be described as a function of the proportion of other-regarding players, and the implied expected values of behavioural variables can be calculated. As an example, in UG when $\rho_o < 0.25$, self-regarding proposers offer 0 and other-regarding proposers offer 10. Then, for a given $\rho_o < 0.25$, the expected offer from the proposer population is $10\rho_o + 0(1 - \rho_o) = 10\rho_o$. Similarly, when $\rho_o > 0.25$ self-regarding proposers offer 5 and other-regarding proposers offer 10, implying an expected offer of $10\rho_o + 5(1 - \rho_o) = 5 + 5\rho_o$. In cases where there are multiple equilibria, each class of equilibria is chosen with equal probability, and if some class permits multiple messages or offers, the arithmetic mean of these

²⁰This assumption may be restrictive if individuals adopt self-serving notions of fairness (e.g., Babcock et al., 1995; Binmore et al., 1998).

Figure 6: Equilibrium behaviour in TG(0.5), conditional on ϕ_0 and ρ_0



is chosen.

The second assumption we make is that since types are unobservable, and there is little justification for any particular assumption about their distribution in the population, we apply a version of Laplace's (1814) Principle of Insufficient Reason, and treat the fraction of other-regarding types as uniform over $[0, 1]$. Then, predicted behaviour in a particular game is simply the unconditional expectation given a uniform distribution over ρ_o .²¹ For example, the expected offer in UG is given by

$$E(x) = \int_0^{0.25} 10\rho_o d\rho_o + \int_{0.25}^1 (5 + 5\rho_o)d\rho_o,$$

which evaluates to approximately 6.406 (just over 32% of the cake). The resulting point predictions for offers and overall acceptances, and for the TG games, messages, over-statement, and the probability of acceptance of an unseen offer, are shown in Table 9.

B English translation of experimental instructions – TG(0.5) treatment

Welcome,

Thank you for your participation. The aim of this study is to understand how people make decisions in certain situations. From now on, talking to each other is prohibited. Please also turn off your mobile phones.

²¹This assumption is somewhat unsatisfactory, since it implies that subjects have common knowledge of the true value of ρ_o , and of which particular equilibrium is selected if multiple equilibria exist, while we the researchers (and likely you the reader) do not know either of these. However, this kind of assumption is commonly made: see, for example, nearly any application of quantal response equilibrium (McKelvey and Palfrey, 1995). Also, any alternative assumptions (e.g., imposing a particular value of ρ_o) seem to be even more difficult to justify. (See Morris, 1995, for an explanation of how this kind of assumption became universal in economics.)

Table 9: Characteristics of equilibria of UG, TG(0.5), TG(0) based on the model

	Expected offer	Expected message	Expected over-statement	Prob(unseen offer accepted)	Overall acceptance probability
UG	32.0	—	—	—	0.974
TG(0.5)	28.7	34.6	5.9	0.900	0.839
TG(0)	25.0	39.0	14.0	0.722	0.722

Note: offers, messages and over-statements expressed as percents of the cake.

If you have a question please raise your hand. We will come to you and answer your question. Please do not hesitate to ask questions, since it is very important that all participants understand the rules in this study. The experiment will be conducted on the computer and you will make all your decisions using the computer. You will earn a monetary reward in the game you will play during the experiment. The amount you will earn depends on your decision and the decisions of other participants. This amount and the participation fee will be paid to you in cash at the end of the experiment.

The Roles and the Groups:

At the beginning of the experiment, you will be randomly assigned to one of these roles: Player A or Player B. This role is fixed.

The game you will play will last for 5 periods and at every period you will be matched in groups of 2 players. In your group, there will be another player whose role is different than yours. Therefore in each group there will be a Player A and a Player B. This matching is randomly determined and you will never be matched with the same person more than once.

You will not learn the identities of people that you're matched with. Similarly, other players who are matched with you will not learn your identity.

At every period you will play the following game:

The Game:

There are 20 TRY to be split between Player A and Player B. Player A proposes how much of this 20 TRY to offer to Player B. At the same time, Player A chooses a message to be sent to Player B. In his message Player A will tell Player B how much his offer is. This message can be accurate or not.

After this, Player B will see Player A's message. With probability 50%, Player B will also see Player A's actual offer, and with probability 50%, Player B will not see Player A's offer.

Player B will then decide whether to accept the offer of Player A or not.

If Player B accepts the offer, the earnings of the players will be determined according to the actual offer of Player A. That is, the earnings of players will be as follows:

Player A: **20 – offer**

Player B: **offer**

If Player B rejects the offer of Player A the earnings of both players will be 0. In either case, Player A's message has no effect on either player's earnings.

At the end of each period, a summary of the current period (offer, message, acceptance decision, earnings) will be presented to both players. We will randomly select one of the 5 periods and your earnings from this part of the experiment will be equal to your earnings in the selected period. Each period is equally likely to be selected, therefore it is in your best interest to decide carefully during each period.

C Questionnaire questions (English translation)

Attitudinal questions (all answered on a scale from 1=absolutely wrong to 6=absolutely right):

q1: I won't hesitate to ask my boss for a pay rise.

q2: I might invest in risky assets.

q3: Lying is difficult for me.

q4: I can choose someone as a roommate without knowing that person well.

q5: I don't hesitate to wear unconventional clothes.

q6: I might argue with a friend who has a very different opinion on an issue.

q7: If I lose my wallet, I believe that the person who finds it will bring it to me with the things in it.

q8: I trust my friends in money issues.

q9: I can spend money without thinking about the consequences.

q10: I can admit it easily when my tastes are different from my friends.

q11: I won't hesitate to move to a different city.

q12: If I would have an unexpected money windfall, the first thing I would do would be to share with people I know.

q13: I can take a job at which I will get paid on commission only.

q14: If I am rich enough, I can lend high amounts of money to my friends.

q15: My verdicts about others' trustworthiness generally turn out to be correct.

Demographic questions:

Age: age of the subject (in years).

Sex: gender of the subject (1=female, 0=male).

Living: living arrangement for the subject (0=student housing, 1=with family, 2= with friends, 3=alone).

Siblings: number of siblings of subject.

Older siblings: number of siblings who are older than the subject.

Major: subject's major (2=economics, 1=other business, 0=other).

Econ: number of economics classes (censored at 4).

D Sample screen-shots from the experiment

Below are sample screen-shots, typical of those that would have been seen in the CG(0.5) cell of the experiment. Screen-shots from the other cells are similar, and can be obtained from the corresponding author upon request.

Proposer decision screen – CG(0.5)

Out of 20 TL, how much do you offer to player B?

In your message to player B, how much out of 20 TL would you tell him that you sent him?

Player B will see your message. He/she, however, will see your actual offer with 50% probability.

Continue

Responder decision – CG(0.5), offer not observed

Based on the random draw of the computer, you will only see the message of Player A

Player A informed you that he/she is offering 10 TL out of 20 TL.

As mentioned before, since at this stage you cannot observe the actual offer of Player A, this message could be accurate/truthful or not.

Would you like to accept the division of 20 TL based on the actual offer of Player A?

No
 Yes

Continue

Responder decision – CG(0.5), offer observed

Based on the random draw of the computer, you will see both the message and the actual offer of Player A

Player A informed you that he/she is offering 10 TL out of 20 TL.
The actual offer of Player A is 7 TL out of 20 TL.
Would you like to accept the division of 20 TL based on the actual offer of Player A?

No
 Yes

Continue