

The Effect of Stake Size in Experimental Bargaining and Distribution Games: A Survey

Emin Karagözoğlu¹ · Ümit Barış Urhan²

© Springer Science+Business Media Dordrecht 2016

Abstract We review the literature on bargaining and distribution experiments to investigate whether changes in stake size have significant effects on behaviour in laboratory/field settings. We conclude that experiments in this field do not lead to clear/common results. The joint presence of opposing factors (e.g., increasing relative risk aversion and increasing cost of fairness) might be one reason contributing to this. Moreover, we argue that variables such as subjects' financial conditions, cognitive abilities, risk attitudes, loss-aversion, justice orientations, and relevant personality characteristics should be controlled in laboratory experiments to understand the effect of stake size on behaviour, more clearly. Finally, quasi-experiments using data from (very) high-stake games/events and meta-analysis studies should complement (individual) controlled experiments.

Keywords Bargaining experiments · Dictator game · Distribution experiments · Fairness · Stake size · Ultimatum game

1 Introduction

1.1 Motivation

In this study we examine the effect of changes in stake-size on behaviour in experimental bargaining and distribution games. The question is closely related to the use of material/financial incentives in experiments and constitutes one of the methodological

✉ Emin Karagözoğlu
karagozoglu@bilkent.edu.tr

¹ Department of Economics, Bilkent University, 06800 Bilkent, Ankara, Turkey

² Department of Economics, Middle East Technical University, 06800 Çankaya, Ankara, Turkey

principles that define experimental economics. The use of material incentives/rewards is a major debate between experimental economists and (experimental) psychologists. There have been many studies on this controversial issue with different results. For example, [Smith and Walker \(1993\)](#), [Camerer and Hogarth \(1999\)](#), [Holt and Laury \(2002\)](#), [Camerer \(2010\)](#), [Murayama et al. \(2010\)](#), and [Kang et al. \(2011\)](#) all found that material rewards have significant effects on decisions and motivations of subjects. [Kühberger \(2001\)](#), on the other hand, found that the use of hypothetical rewards would be sufficient in decision-making experiments, and [Read \(2005\)](#) argued that monetary incentives are neither necessary nor sufficient to generate cognitive exertion, motivational focus, or emotional triggers.

If one acknowledges the effect of monetary incentive mechanisms on subjects' behaviour, this brings yet another important question: does stake-size affect behaviour? This is the main question we deal with in this survey study. The significance of the answer to this question, which concerns both economic theorists and experimental economists, can be summarized under two main titles: (i) the generality of economic theories and (ii) the external validity of experimental studies.¹ Needless to say, our conclusions would also be of interest to (experimental) psychologists.

In most (mainstream) theoretical models in economics, results (or predictions) are independent of the amount of monetary stakes. For example, in the subgame perfect Nash equilibrium of the ultimatum game with *standard* preferences for players, predictions about players' behaviour are independent of the amount of resources distributed in the game. Similarly, the standard model of *rational* behaviour would not predict agent's behaviour in the dictator game to change even if the money at stake changes. Finally, the scale invariance axiom (and its variants) frequently utilized in axiomatic bargaining theory leads to similar results as far as stake-size effect is concerned. Therefore, experimental studies that vary monetary incentives can play an important role in testing the generality or robustness of these theoretical models. Hence, by examining the results of various bargaining and distribution experiments in which different stake sizes are used, our study may help us understand the robustness of theoretical models in this field.²

On the other hand, one of the main criticisms against experimental economics is that monetary incentives used in experiments are quite small, thereby questioning the external validity of the results (see [Bolte 1990](#); [Falk and Fehr 2003](#), p. 401; [Guala 2005](#)). [Levitt and List \(2007\)](#) took an eclectic position against this criticism by arguing that experiments using small monetary incentives are still important since they can facilitate a better understanding of our daily economic transactions. However, they also pointed out that it is necessary to use larger monetary incentives to test whether it makes a difference, and accordingly, to provide feedback on whether economic agents change their behaviour as stakes change. Analysing the results of experiments that use

¹ A relatively recent yet influential book by [Bardsley et al. \(2010\)](#) devoted two chapters to these issues.

² We do not focus on the effects of a presence (real) versus absence (hypothetical) of financial incentives here (see [Hertwig and Ortmann 2001](#) for a great discussion). Nevertheless, for the sake of giving a more complete picture we will include some studies comparing behavioral patterns in ultimatum, dictator, and trust games with real versus hypothetical stakes.

different stake sizes offers an opportunity to investigate whether criticisms concerning the external validity of experiments are empirically valid.

These results would also be important for more pragmatic reasons. Firstly, if there is enough evidence in favour of *no stake size effect hypothesis*, this can provide a basis for running (bargaining and distribution) experiments at more reasonable costs (see [Gillis and Hettler 2007](#) for a similar argument). Secondly, if the accumulated evidence suggests that small variations in stake-size cannot change behaviour, then researchers focusing on questions other than stake-size but varying the stake-size (at a small margin) for the sake of robustness, should either abandon these efforts or start using much larger stake-size variations.

Finally, a word about our focus on bargaining and distribution experiments. [Camerer and Hogarth \(1999\)](#) found that stake-size effects may vary across different tasks in a systematic fashion (see [Hertwig and Ortmann 2001](#) for a similar result). Therefore, we do not claim that our conclusions extend to all possible tasks, economic environments, games etc. We limit our attention to bargaining and distribution games. Since the emergence of experimental economics, bargaining and distribution games have always attracted great interest. It would possibly not be an exaggeration to say that after individual choice under uncertainty, bargaining and distribution games are two of the most-studied settings in the experimental economics literature. Not surprisingly, ultimatum and dictator games have been very popular test-beds for stake-size effects.

1.2 Earlier Studies

There are a few earlier studies which are similar to the study presented in this paper. In the first, [Smith and Walker \(1993\)](#) analysed 31 studies—using a variety of stakes and monetary incentives—in light of a formal model which assumed that decision-making has a *cognitive cost*. They discovered that for higher stakes, the results centre around the predictions of theoretical models, i.e. an increase in monetary incentives forces/encourages individuals to behave more *rationally*. Another conclusion of the study was that an increase in monetary incentives decreases the variance observed in the behaviour. [Camerer and Hogarth \(1999\)](#) analysed 74 experimental studies using (i) no monetary incentives, (ii) low monetary incentives, and (iii) high monetary incentives. They investigated the effects of monetary incentives on behaviour in various experimental games. They found that, especially in decision-making experiments in which extra effort had a positive impact on the performance, higher monetary incentives increased performance.

In his book, [Camerer \(2003\)](#) included a subsection on stake-size effects in ultimatum game and bargaining experiments. Based on findings from earlier studies on the presence of stake-size effects, he concluded that if there is a stake-size effect at all, it is surprisingly weak.³ [Konow \(2005\)](#), in a survey paper, studied how information and

³ [Bearden \(2001\)](#) surveyed ultimatum game experiments and he had a section in the survey on the stake size effect. He concluded similarly by saying that the effects found were weaker than he thought and further data is needed for better understanding.

stake size influence subjects' fairness concerns and behaviour in distribution/sharing experiments. The conclusion he arrived at by considering a number of ultimatum and dictator game experiments was that neither subjects' behaviour nor their fairness concerns exhibit systematic variations across experiments using low stakes and high stakes.

Our paper differs from the first two due to our focus on bargaining and distribution experiments, and it departs from all four due to our recommendations for subject recruitment and data collection. Finally, we also cover the most recent studies, some of which used considerably larger stakes and had results drastically different than the earlier ones.⁴

1.3 Conclusions and the Organisation of the Paper

In light of our observations, we argue that (i) the number of papers studying the effect of stake-size on individual behaviour is not sufficient to reach solid conclusions (possibly except for ultimatum games), (ii) including ultimatum game experiments, findings do not deliver a clear and consistent message as to the effect of stake-size on individual behaviour, (iii) differences in findings and the lack of a unique message may stem from the co-existence of effects such as increasing relative risk aversion and increasing fairness costs, which are usually not controlled in analyses,⁵ small sample sizes, uncontrolled yet relevant factors, different samples (e.g., students, professionals), and different incentive mechanisms (e.g., real vs. hypothetical stakes, full vs. lottery payments etc.) used in different studies (iv) liberally high stake-sizes may be needed (e.g. [Andersen et al. 2011](#); [Goldreich and Pomorski 2011](#)) to be able to observe behaviour that is in line with predictions of theoretical models in bargaining and distribution games experiments, and (v) quasi-experiments using data from high-stake tournaments, contests, television shows etc. and meta-analysis studies are two promising (and complementary) sources in our quest for answers.

One particular message that these observations do give is that experimental studies—on the effect of stake-size on behaviour or in general—require more “control”: it is essential to conduct experiments that control subjects' income levels, cognitive abilities, financial needs, risk attitudes, loss-aversion, fairness concerns, and other relevant personality traits to thoroughly understand the effect of stake size on individual behaviour. The presence of uncontrolled yet relevant factors and an exclusive focus on *average* behaviour is likely to be a reason for mixed results.

We also want to emphasize that the type of strategic interaction, the payment method, and the particular experimental design used in a study may have an impact on

⁴ It is worth mentioning that we included also working papers that are not published yet to avoid possible publication biases that may arise in survey studies. We also want to mention that we cover around fifty studies after 1999, which is the publication date of [Camerer and Hogarth \(1999\)](#).

⁵ For instance, a proposer who becomes increasingly risk averse as the stake increases may tend to make proportionally more generous offers. On the other hand, if he believes that fairness is a normal good (for the responder) and the cost of behaving fair-mindedly increases as offers become more generous in absolute terms, he may make more generous offers in absolute terms but less generous in relative terms. Hence, the net effect can be indeterminate.

the stake-size effect. In what follows, we report studies that *marginally* vary the stake size (e.g., \$1–\$6 or €4 vs. €8) and find a significant stake-size effect and studies that significantly vary the stake size (e.g., \$10 vs. \$100, or \$200 vs. \$20,000) and do not find any. We think that understanding and pointing out the characteristics of strategic interactions and design elements that have a systematic influence on stake-size effects is a promising research agenda.

Another point we want to draw attention to is about the use of quasi-experiments. Certain real-life contests, games, television shows etc. are potentially good test-beds for stake-size effects since they involve very large sums of money which are practically impossible to have in lab experiments. We think that an increase in the number of such studies will be beneficial. We also believe that lab experiments that use subject recruitment and data collection methods we propose in this paper and quasi-experiments of the sort we mention above are complements rather than substitutes. The former aims to (and can) have a lot of control, but cannot afford to have very large stakes; and the latter has a certain lack of control (compared to lab experiments) but can afford very large stakes.

Finally, more meta-analysis studies would be helpful in obtaining more robust and statistically powerful results than individual experiments. They also offer more objective and informative results than literature reviews.

The paper is organized as follows: in Sect. 2 we review the literature to find theoretical models/arguments for the presence of stake-size effects, mention predictions of these models, and outline the essential factors that should be incorporated in theoretical models that aim to study/capture stake-size effects. In Sects. 3 and 4 we review bargaining and distribution experiments, respectively. In these sections, we follow a chronological order in presenting the results from different studies. In the “Appendix” we also provide a summary table (see Tables 5, 6) where we group these studies according to their findings (e.g., whether a stake-size effect is found or not). We provide some methodological suggestions for subject recruitment and data collection in Sect. 5. Section 6 concludes.⁶

2 Theoretical Arguments

We believe that going through theoretical arguments about stake-size effects is almost a pre-requisite for studying the experimental literature on the topic since doing so can help in setting benchmarks about what to expect. Accordingly, we first touch upon theoretical arguments dealing with changes in stake size. For reader-friendliness (e.g., to avoid a lot of notation), we stick to verbal argumentation of theoretical models. The summary of predictions from these studies can also be found in Table 1.

Variations of standard theoretical models already set forth some predictions about how stake size affects individual behaviour. For example, under social preferences or the possibility that players can make mistakes, Markowitz’s (1952) and Menezes and

⁶ In some of the studies we investigate, the effect of stake-size on individual behaviour is not the main research question. Rather, the authors implement different stake-sizes to test the robustness of their results. When we report such studies, we focus on those parts where they test for stake-size effects.

Table 1 Summary table of theoretical predictions

Reference	Argument/theory	Implication or prediction
Markowitz (1952)	Increasing relative risk aversion	Proposals, in an ultimatum game, increase with stake-size
Menezes and Hanson (1970)	Increasing relative risk aversion	Proposals, in an ultimatum game, increase with stake-size
Kahneman and Tversky (1979)	Framing/loss-aversion	Ultimatum proposals increase with stake-size
Thaler (1988)		Responders' minimum acceptable offers in an ultimatum game increase less than proportionally as stakes increase
Bolton (1991)	Comparative model of bargaining	If fairness is a normal good, proposals in an ultimatum game increases with stake-size
Telser (1993)	Cost of fairness	Fairness is a normal good. Hence, as the cost of rejecting an offer increases, rejections will be less frequently observed
Smith and Walker (1993)	Smith–Walker hypothesis	As the money at stake increases, people will make better, more rational decisions and these decisions will be around rational predictions with a smaller variance
Rabin (1993)	Fairness equilibrium	Material concerns dominate fairness related concerns as stakes increase significantly. As a result, reciprocal behaviour will be observed less under very high stakes
Hoffman et al. (1996)		The tendency to avoid the rejection risk in the ultimatum game increases with the stake-size
Bolton and Ockenfels (2000)	Theory of equity, reciprocity, and competition	Rejections of unfair offers will vanish as stake size increases
Dickinson (2000)	Rabin's reciprocal kindness theory	Ultimatum offers become less fair and less fair offers are accepted more as stake size increases
Bolton and DeBruyn (2008)		Agents make less mistakes and their behaviour is less erratic around rational decisions
Camerer et al. (2015)	Revelation principle of Myerson (1979, 1984)	Disagreement rates in an unstructured bargaining with asymmetric information on the stake-size will be monotonically decreasing in stake size

Hanson's (1970) *increasing partial risk-aversion hypothesis* suggest that when the amount to share is high, subjects who make offers (as the party facing the risk of being rejected) will make proportionally more generous offers compared to the cases when

the amount to share is lower in the ultimatum game.⁷ If the behaviour of the player who faces risk is in line with the increasing partial risk-aversion hypothesis, then his tendency to avoid risk (of rejection) increases as the money at stake increases. Thus, he will make proportionally more generous offers (for instance, given that the responder has social preferences). The study by Hoffman et al. (1996) featured a similar argument where, again, the tendency to avoid risk (of rejection) increases with stake-size.

Another prediction on how stake-size influences individual behaviour can be obtained using the well-known *framing effect* by Kahneman and Tversky (1979). According to this argument, individuals tend to base their decisions more on potential losses than gains. Therefore, if the proposer believes that the responder may value fairness, then the proposer's decision, when making an offer in the ultimatum game, will depend more on the amount that he may lose in case of a rejection rather than the amount that he may gain in case of an acceptance. As a result, proposers will tend to make more generous offers in games with high stakes since the amount lost is larger in case of a rejection.

Thaler (1988), observing the first ever ultimatum game experiment in the literature, made some predictions about an ultimatum game with \$1000 or \$100,000. His prediction was that the minimum amount that responders are willing to accept will increase non-linearly—at a decreasing rate—with the stake size. Additionally, he argued that the minimum acceptable offer is proportional to responder's wealth; and he concludes that resisting unjust offers (behaving fair-mindedly) is a *normal good*. Note that all three arguments mentioned above require deviations from standard (e.g., selfish, rational) preferences.

Thaler's argument overlaps with Telser's (1993) *cost of fairness* argument. According to Telser (1993), there is a *cost* associated with behaving in line with fairness considerations. For instance, the amount of money a responder foregoes by rejecting an unfair offer in an ultimatum game can be seen as the cost of behaving fair-mindedly. If behaving fair-mindedly is a normal good (see Zwick and Chen 1999 for supporting results), then other things remaining same, the individual is expected to behave less fair-mindedly as the price (or the cost) of doing so increases. The impact of this on the ultimatum game behaviour is that responders will be likely to accept offers that are high in absolute terms but low in relative terms (e.g., \$1000 offer in an ultimatum game where \$100,000 is at stake).

Bolton (1991), in a seminal piece of work, argued that agents may derive utility not only from absolute monetary earnings but also from relative monetary earnings. What he called as *relative money* is the difference between what one agent gets from bargaining agreement and what the other agent gets. In his model, the effect of increasing stake-size depends on whether fairness is a normal or an inferior good and subjects' risk attitudes. In particular, he argued that if fairness is a normal good, then offers will increase as stake-size increases under complete information.

Almost all the predictions presented above are related to individual behaviour in ultimatum games. On the other hand, Smith and Walker's (1993) study is related to how individuals react behaviourally to increases in financial/monetary incentives. In

⁷ This hypothesis is empirically supported by studies such as Binswanger (1981), Kachelmeier and Shehata (1992), Bombardini and Trebbi (2005), Andersen et al. (2008), Baltussen et al. (2008) and Post et al. (2008).

line with the argument that claims decision-making is an activity which has a *cognitive cost*, [Smith and Walker \(1993\)](#) proposed that higher potential losses or gains will lead individuals to think more thoroughly and make more rational decisions.⁸ Therefore, [Smith and Walker \(1993\)](#) predicted that subject behaviour in experiments where large financial incentives are used will be accumulated around the predictions of theoretical models which assume rationality. Analyses they conducted on 31 experimental studies confirmed their predictions.

[Rabin \(1993\)](#) introduced the *fairness equilibrium* in an attempt to incorporate fairness into game theory and economics. In his model, reciprocating opponent's actions would enhance agent's *fairness-related payoff*. Thus, positive and negative reciprocation both have a potential to increase agent's payoffs. The crucial element is that there is a trade-off between fairness payoff and material payoff; and fairness payoff is bounded whereas material payoff is not. Rabin's model predicts that if stakes increase significantly, material payoff concerns will dominate fairness related concerns. As a result, less reciprocity should be observed in games with considerably high stakes; and predictions of the standard theoretical models with selfish and rational agents should be consistent with data if stakes are liberally high. Later, [Dickinson \(2000\)](#) applied Rabin's reciprocal kindness theory to an ultimatum bargaining model and derived qualitatively similar predictions for proposer offers and responder decisions.

In an elegant calibration exercise that used data from multiple bargaining studies [Bolton and DeBruyn \(2008\)](#) outlined a theoretical model of bargaining behaviour based on quantal responses. Therefore, they included in their model the possibility that agents' make mistakes.⁹ Their model predicted that as the stake-size gets larger, agents will make less mistakes and their behaviour will be less erratic around rational decisions. In the same paper, the authors showed that this model fits the data well.

Recently, [Camerer et al. \(2015\)](#) took on a challenging task. Unstructured bargaining experiments mostly use cooperative bargaining solution concepts to derive theoretical predictions since the strategy spaces of such bargaining protocols are extremely complex to be modelled as a non-cooperative game. Nevertheless, their model can provide predictions on disagreement rates independent of the bargaining protocol, and the results are only based on information conditions. They modelled an unstructured bargaining game where the pie size is known by only one of the players and the other player only knows the distribution from which the pie is drawn. They showed that if the bargaining mechanism satisfies *individual rationality* (both players prefer to participate in the mechanism) and *incentive compatibility* (the informed party has an incentive to tell the true pie size), then (i) the disagreement rates will be monotonically decreasing in the pie size and (ii) the uninformed party's payoff is weakly increasing in the pie size.

⁸ This argument (i.e., decision-making has cognitive costs) is supported by recent neuro-scientific research. For further reading, the reader is referred to [McGuire and Botvinick \(2010\)](#).

⁹ Their model used [Bolton and Ockenfels \(2000\)](#) *inequity averse* preferences as a primitive. Note that, [Bolton and Ockenfels \(2000\)](#) and [Rabin \(1993\)](#) both imply that as the stakes increase, rejections of unfair offers will vanish.

It is worthwhile emphasizing the differences between the two main transmission channels for stake-size effects observed in these studies. In particular, models with standard preferences and full rationality cannot account for stake-size effects. Incorporating the possibility of mistakes or fairness concerns into theoretical models opens a transmission channel for stake-size effects on behaviour. Naturally, predictions from such models will be different than that of models with standard assumptions. In particular, in models where players can make mistakes, increases in stake-size will decrease the likelihood of such mistakes and bring equilibrium behaviour closer to the predictions of models with fully rational players. On the other hand, in models where players have social preferences, most researchers found that increases in stake-size leads to more selfish behaviour.

Setting up a full-fledge theoretical model of how individual behaviour responds to variations in stake size for all the (non)strategic distribution problems we cover goes beyond the aims and scope this paper. However, we can still outline some of the elements such a model should have.

1. *Decision Quality* Strategic decisions involve/require significant cognitive activity and such activity is costly (at best, cognitive resources are limited). Thus, a *good* model of stake-size effects should incorporate decision quality or cognitive effort as a function of (expected) return. Allowing mistakes and assuming that the likelihood of mistakes is decreasing in money at stake, a la quantal response models would be one way to do so.
2. *Risk Attitudes* Most of the situations/problems we cover here involve risk for one or more decision-maker(s). Increasing relative risk aversion appears to be an essential element to include if one wants to explain behavioural differences across low versus high stake strategic situations involving risk.
3. *Prospect Theory, Reference Points, Loss Aversion* The presence of prospect theoretic preferences and/or reference-dependent behaviour can influence the way monetary payoffs are perceived/evaluated. Moreover, recent studies (e.g., [Fehr-Duda et al. 2010](#); [Vieider 2012](#)) showed that stake-size variations can also influence the way individuals weigh probabilities on good and bad outcomes. If there are obvious reference points or clear reasons to believe in reference-dependent behaviour in the economic environment to be modelled, incorporating these elements may be necessary in order to study stake-size effects.
4. *Non-monetary Factors Influencing Utility* Experimental evidence showing that individuals care about fairness, equality (or equity), honesty, social norms, loyalty, keeping promises etc. is abundant. Moreover, the relative price ratio between any of these concerns and monetary payoffs is most probably not constant but rather depends on the money at stake. Furthermore, neither increasing relative risk aversion nor prospect theory preferences alone can imply stake size effects if individuals have standard, selfish preferences with an exclusive focus over monetary payoffs. Therefore, depending on the specific economic environment, one or more of these factors (e.g., concerns for fairness, honesty, promise-keeping, loyalty etc.) should be incorporated into a model that aims to capture stake-size effects. One way to do this may be to allow inequity aversion parameters in Fehr–Schmidt

(1999) (or Bolton and Ockenfels 2000) models to be functions of stake size rather than constants (see Güth et al. 2007 for a similar argument).

5. *The Source of the Money at Stake* The experimental evidence since early 2000s consistently shows that bargaining and distributive behaviour differs in games where the money at stake is produced by the involved parties from the ones where the money is *mana from heaven* (see Karagözoğlu 2012). In particular, if and when the resource (or money) is jointly produced through efforts, investments or any costly actions, then they play a significant role in the distribution stage. A typical finding is that individuals (more strongly) defend their *rights or entitlements* which they derive from their contribution to the production and an accompanying justice ideal (e.g., egalitarianism, liberalism, libertarianism). Therefore, stake-size variations caused by individuals' efforts may have different impacts on behaviour than variations caused by exogenous/stochastic factors.

Needless to mention, some of these factors are more essential in certain economic environments than in others. For instance, resorting to non-monetary factors influencing utility would be more reasonable than to risk-aversion or decision quality if one studies behaviour in dictator games.

3 Bargaining Experiments

In this section, we review the studies on bargaining experiments where stake size is varied. We group bargaining experiments according to the bargaining protocols they employ. The most popular bargaining protocol in the experimental literature is the ultimatum game. Accordingly, we first focus on ultimatum game experiments. Some other bargaining protocols used in bargaining experiments are Nash demand game, unstructured/semi-structured bargaining, and Rubinstein–Ståhl (or alternating offers) bargaining. We group them as “other bargaining games”.

3.1 Ultimatum Games

Take two agents A and B who are anonymous to each other. Assume that a certain amount of money is given to agent A (proposer) who is then asked to decide on what part of this money to offer to agent B (responder). Denoting the amount of money with M , let's say agent A offers X to agent B and gets $M - X$ for himself. Now, agent B has two options: accept A 's offer or reject it. If he accepts, M will be distributed in line with A 's offer, i.e., A receives $M - X$ and B receives X . If he rejects, both agents receive zero.

The simple bargaining game we describe above is known as the *ultimatum game*. Under standard rationality assumptions and standard preferences, the theoretical prediction for the outcome of this game is: A offers the lowest possible amount to B and B accepts this offer.¹⁰ The reason is: if both sides are rational (and if the rationality

¹⁰ This is the subgame perfect Nash equilibrium (SPNE) prediction. Nash equilibrium (NE) concept predicts other outcomes, as well. Nevertheless, SPNE is much more frequently employed in extensive form games (such as the ultimatum game) due to a weakness of NE in capturing sequential rationality.

of agents are common knowledge), then it is common knowledge that B will accept anything greater than zero. Thus, A will offer the lowest possible amount.

The first experimental study testing this prediction was conducted by Güth et al. (1982). This study featured two main experiments (simple vs. complex game). The purpose of these experiments was to investigate how the level of complexity affects bargaining results. This classic study is one of the first to show that individuals may systematically behave differently than the predictions of (standard) game theoretic models. In particular, in the experiments of Güth et al. (1982), proposers offered noteworthy amounts and responders generally did not accept low offers.

With a different perspective, Thaler (1988) brought up the question: How would proposals and responses look like when the stake size is much larger (e.g. \$1000 or \$100,000). First (partial) answer to this question was given in Camerer's work in 1990. In his study, Camerer (1990) showed that varying the stake size (\$12 and \$120) does not lead to a significantly different behaviour: proposals (respectively, 39 and 38 %) and accepted lowest amounts (respectively, 21 and 15 %) were identical across two stake sizes.

The results of early ultimatum game experiments were very interesting as much as they were in stark contrast with what theory suggests. A very natural question that arose was the generalizability of these results across different cultures. Roth et al. (1991) analysed behaviour in the ultimatum game with experiments conducted in Israel (Jerusalem), Japan (Tokyo), USA (Pittsburgh) and Yugoslavia (Ljubljana).¹¹ Throughout these experiments, each subject played the same game 10 times. In the experimental sessions in the USA, the authors used \$10 versus \$30. They did not find any difference in behaviour across ultimatum game experiments with \$10 and \$30.

A decade after Güth et al. (1982), one popular explanation for the behaviour observed in ultimatum game experiments all around the world was fairness. Responders may reject small but positive offers because they consider them to be unfair; and proposers, either being fair-minded people or being afraid of facing a fair-minded opponent, propose fair shares. Forsythe et al. (1994), in their seminal work, examined whether the ultimatum proposals are related to proposers' fairness perceptions. They argued that if the main factor influencing proposals is fairness, then proposer behaviour in ultimatum and dictator games¹² should not differ. Their study tested the *fairness hypothesis* by comparing proposer behaviour across ultimatum and dictator games. Another tested hypothesis was whether paying subjects has an effect on their decisions or not. In addition, the study also analysed the effect of increasing the value at stake.

The experiment included both ultimatum and dictator games with \$5 stakes. To test the empirical validity of their hypotheses, two experiments were carried out. To observe the effect of the change in stake size, another experiment with \$10 stake was conducted. Results did not show any significant difference in behaviour across experiments with

¹¹ This article consists of two sections which are the bargaining game and the market game. From the point of the current study, only the bargaining section is explained.

¹² In the dictator game, like in the ultimatum game, first player proposes a division. However, there is no decision to be made for the second player, i.e., the second player cannot accept or reject. In other words, the first player rather dictates the division. Hence the name, dictator game.

\$5 and \$10 stakes. On the other hand, when stakes were hypothetical (i.e., no real monetary incentives) proposers behaved more fairly.¹³ More importantly, dictator and ultimatum game behaviour of proposers were found to be different, rejecting the explanatory power of fairness hypothesis: in the dictator game experiment with \$5 stake, 36% of dictators offered nothing, while only 21% made an equal or better offers. However, in the ultimatum game experiment with the same stake size, 65% of the offers were equal or better offers. The same result was valid for \$10 stake: the percentage of equal offers in the ultimatum game increased up to 75%, whereas in the dictator game and the percentage of zero offers is 21% and the percentage of equal or better offers dropped to 21%. This was a very striking result at the time but in light of the literature we survey in this paper, it is not very surprising given that stakes used in the study were both on the lower end.

Tompkinson and Bethwaite (1995) re-investigated Thaler's (1988) question on the effect of higher stake size on proposals and responses in the ultimatum game. The difference between this study and many others in the literature lies in their incentive structure and the subject profile: subjects in this study were professional lawyers and they were asked to share \$10 or \$10,000, hypothetically. They were also asked their minimum acceptable offers. Thus, each subject took the role of both proposer and responder. The experimental design included three different games:

- (i) *Game 1* The subjects were asked how much they would offer as an allocator and specify their minimum acceptable offers. It's a \$10 game and offers can be multiples of \$1. Finally, the minimum possible offer is \$1.
- (ii) *Game 2* The same with Game 1. The only difference is that it's a \$10,000 game and offers can be multiples of \$1. The minimum possible offer is \$1.
- (iii) *Game 3* The same with Game 2. The only difference is that offers can only be multiples of \$1000. The minimum possible offer is \$1000.

Results, as can be seen in Table 2, show that the average proposal was \$4.3 and the average accepted offer was \$3.6 in Game 1. In Game 2, these numbers turned out to be \$3611 and \$2599, respectively. Furthermore, in Game 3 they slightly increased to \$3790 and \$2883, respectively. In Game 1, 29% of the subjects were willing to accept the minimum possible offer (i.e., \$1). In Game 2, only 16% of the subjects were willing to accept the minimum possible offer (i.e., \$1). Understandably, in Game 3, 44% of subjects were willing to accept the minimum possible offer (i.e., \$1000). Looking at modal minimum acceptable offers across three games tell a similar story: the modal minimum acceptable offers are 5, 5000, and 1000 in Game 1, 2, and 3, respectively. Mean offers dropped as stake-size increased. Finally, the percentage of subjects applying a fair-division (game theory) strategy of "accept half of the pie, offer half of the pie" ("accept and offer the minimum possible offer") is lower (higher) in Games 2 and 3 (Game 3) than in Game 1 (Games 1 and 2).

These results are in line with the hypothesis that increasing the stake and denominations both lead to more selfish proposals. This is one of the earliest studies that

¹³ Later, Gillis and Hettler (2007) conducted a multiple-round ultimatum game experiment with real versus hypothetical stakes (\$10) and found mixed results. In particular, subjects offered—on average—less in the hypothetical stake treatment. However, this effect was not present in the first round of play but became apparent in later rounds.

Table 2 Tompkinson and Bethwaite 1995—summary of results

	Game 1		Game 2		Game 3	
	Offers	Min. accept.	Offers	Min. accept.	Offers	Min. accept.
Mean	4.3	3.6	3611.7	2599	3790.7	2883.7
Median	5	5	5000	3000	5000	3000
Mode	5	5	5000	5000	5000	1000

This table is regenerated from Table 1 in Tompkinson and Bethwaite (1995)

found a stake-size effect. Nevertheless, the results relied on hypothetical decisions. Given the results of the studies showing that monetary incentives influence behaviour, this raises questions on their generalizability.

All studies investigating proposer and responder behaviour in the ultimatum game and varying stake size have assumed that both players have complete and symmetric information about the size of the stake. Straub and Murnighan (1995) is an exception: they asked what would happen when subjects have imperfect information on stake size.¹⁴ To answer this question, they asked all subjects the following questions: (i) if you were a proposer and the other side knew the stake size, what offer would you make? (ii) if you were a proposer and the other side did not know the stake size, what offer would you make? (iii) if you were a responder and you knew the stake size, how would you respond to the offer made to you? (iv) if you were a responder and you did not know the stake size, how would you respond to the offer made to you?

Subjects were presented a series of offers and asked which of them they would accept; and if they would accept none of these offers, they were asked their minimum acceptable offers. First, in the case where stake size was unknown, subjects were given the following offers: \$1, \$0.5, \$0.15, \$0.25, \$5, \$0.01, \$3.5 and, \$0.05. After that, they were given the information that stake size was \$10 and were asked to decide on following offers: \$0.05, \$1.5, \$0.75, \$2, \$0.10, \$0.40, \$5, \$0.01, \$3. After these two questions, for five different stake sizes (\$30, \$50, \$60, \$80, and \$100), they were asked their minimum acceptable offers. Then, they were asked their minimum acceptable offers on two hypothetical stakes (\$1000 and \$1,000,000).

In the following part of the experiment, subjects were confronted with a series of questions about their offers in the presence of different stake sizes. First, assuming that the other agent knew the stake size, they were asked their offers in the following stake sizes: \$10, \$1, 5\$, \$50, \$60, \$80. Additionally, they were also faced with hypothetical stake sizes of \$1000 and \$1,000,000. After this part, they were asked the same questions but now under the assumption that the other agent did not know the stake size. Subjects were paid on the basis of one randomly selected non-hypothetical decision.

When the stake size was unknown, 29 subjects out of 45 declared \$0.01 as their minimum acceptable offer. They even mentioned that they will accept “any offer”.

¹⁴ Another well-known study on ultimatum games where the stake-size is asymmetric information is Nagel and Mitzkewitz (1993). In that study, possible stake sizes were DM1.20, DM2.40, DM3.60, DM4.80, DM6, and DM7.20. The authors reported differences in behaviour across (two versions of the) ultimatum games with different stake sizes.

Nevertheless, when stake-size information was complete, this rate fell to 11 out of 45. Interestingly, the same kind of difference was valid for hypothetical values as well. It can be said that the game theoretical prediction for the ultimatum game under complete and symmetric information on the stake size was observed in the ultimatum game with asymmetric information on the stake size.

Moreover, the authors found evidence for a relationship between the minimum acceptable offer and stake size. While for \$10–\$100 stakes, 20% was the mean minimum acceptable offer, this rate falls to 16.6% for a \$1000 stake and to 10.5% for a \$1,000,000 stake. These results are in line with “fairness” being a normal good. On the other hand, in both perfect and imperfect information cases, offers increased (in the absolute sense) with stake size.

In early studies of ultimatum games testing stake size effects, experiments using stakes higher than \$10 were either done with hypothetical stakes or the random payment method. Hoffman et al. (1996) took one step forward and raised the stakes to (real, not hypothetical) \$100 to test the same question over again. They also included a game with \$10 to compare the results. The experiment implemented two selection criteria in determining subjects’ roles. One of them was the standard random selection whereas the other one determined subject roles on the basis of their performance in a general knowledge quiz. Subjects were asked ten general knowledge questions and according to their results, the first-ranked player was paired with the seventh-ranked, the second-ranked player was paired with eight etc. Here, each better performing player was labelled as a “seller” and his/her partner was labelled as a “buyer”. The ultimatum game was played between the matched pairs. Offers could be multiples of \$10 from a \$100 stake. After responders responded to their offers, payments were made accordingly. The results did not show any significant difference between offers in games with different stakes. This was valid for both selection criteria. However, offers were lower (less generous), again, under both stake sizes when the subject roles were determined on the basis of performances in the quiz. This is in line with a hypothesis that when proposers believe they deserved or earned their role, they will derive entitlements and will offer less. The fact that the authors did not observe any stake-size effect was a significant result in the light of a reasonably large stake of \$100.

Budgetary restrictions are great obstacles to conducting experiments to test the presence/absence of a stake-size effect. Since earlier studies have already shown that mild increases in stakes are not likely to effect behaviour, much larger stakes should be employed. This is particularly difficult in the developed west. Bob Slonim and Roth (1998) found a way to tackle this problem: they conducted their ultimatum game experiment in Slovakia, which enabled them to use very large stakes. Beside the stake-size effect, they also investigated the existence of a *learning effect* with a multiple-round design. Stakes used in the experiment were 60 Slovak Korún, 300 Slovak Korún, and 1500 Slovak Korún. These amounts were equivalent to 2.5, 12.5 and 62.5 hourly wages, respectively. All subjects were randomly divided into proposer and responder roles, and played the ultimatum game ten times, each time with a different player. Subjects started first round without any experience which they gained in the subsequent rounds. All games were played with 1000 points stake size, where offers could be multiples of 5 points. On the other hand, the exchange rate between points and

Korún varied: 1000 points were made equal to 60 Slovak Korún, 300 Slovak Korún, and 1500 Slovak Korún.

Their results showed that when players were inexperienced, there was no difference in their behaviour across games with different stakes. Similarly, without experience, responses to these offers were the same no matter the stake size. Nevertheless, when the game was played repetitively, behaviour started to change in a systematic fashion. In particular, as players became more experienced, the average offers slowly declined in high stakes but not in low stakes; and rejections were lower in high stakes than in low stakes. In particular, over all offers, the rejection rate was 17.1% in low stakes, 12.1% in medium stakes, and 8.8% in high stakes. For offers that gave the responder less than half of the pie, these numbers were 25.6, 16, and 13.6%, respectively.

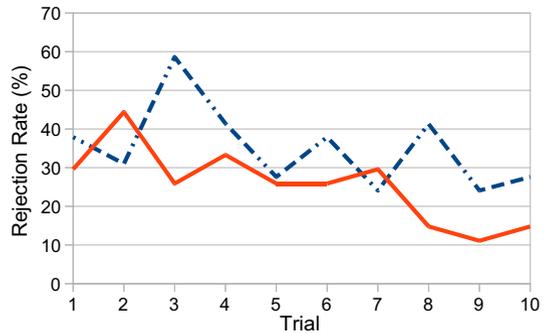
Slonim–Roth study is a seminal contribution for mainly three reasons: (i) it is one of the first studies that showed that rejection rates decrease as stake size increases, (ii) it is the first study to show that experience mediates the effect of stake size on behaviour, and (iii) it shows that the reason for the change in the proposer behaviour may be traced back to the change in the responder behaviour in repeated settings (i.e., proposers learning to make lower offers).

Hoffman et al. (1996) raised the bar by using a \$100 stake, and Slonim and Roth (1998) went further by finding an alternative way (running the experiment in countries where average per-capita income is significantly lower than in the developed west) to run high-stakes experiments. Cameron's (1999) study is along the same fashion with Slonim and Roth (1998): Cameron (1999) ran high-stakes ultimatum game experiments in Indonesia. The study included Rp 5000, Rp 40,000, and Rp 200,000 stakes. These were local currency equivalents of \$2.5, \$20, and \$100, respectively. The experimental design included four different games. Each game consisted of two rounds. Randomly appointed roles of subjects were same in both rounds. Stake sizes in four games for their two rounds was determined as follows: (i) Rp 5000–Rp 5000, (ii) Rp 5000–Rp 40,000, (iii) Rp 5000–Rp 200,000, and (iv) Rp 5000–Rp 200,000. While the first three were real stakes, the fourth was hypothetical. Accordingly, in addition to stake-size effects, both learning effects between the two rounds and the difference between real and hypothetical stakes were studied.

Results showed that the first round offers were similar to offers observed in earlier experiments: on average, 43% of the pie was offered. In the second round, it was observed that there have been more fair offers in all games. Mostly 50–50% divisions were offered. Statistically, there was no difference between the first and the second rounds, and this showed that increases in the stake size did not decrease offers. Proposer behaviour did not converge to subgame perfect Nash equilibrium with an increasing stake size either. All comparisons point towards the same conclusion: proposer behaviour was invariant to stake-size. Moreover, comparing behaviour across real and hypothetical stakes, there was no statistically significant difference between offers. Nevertheless, responders appeared to be less likely to accept low offers when hypothetical offers were present. Finally, a comparison between the first three games showed that as stake size increases, the acceptance rate increases as well. All in all, one can say that these results on the stake-size effect present a mixed picture.

Rabin's (1993) *fairness equilibrium* and *reciprocal kindness* theory soon became very popular among experimental and behavioural economists studying distribution

Fig. 1 Rejection rates across trials (*dashed line* low-stake). This figure is regenerated from List and Cherry (2000) by using the data in Table 2



problems. Dickinson (2000) applied Rabin's *reciprocal kindness* model to ultimatum bargaining to test this theory. As a part of this test, the author varied the stake size from \$1 to \$15 (the intermediate values were \$2, \$4, and \$7). In line with Rabin (1993), the model predicted less kindness being offered by the proposers and less kindness being accepted by the responders as stake size increases. The results based on data from 100 ultimatum games showed that stake size had only a marginally significant influence on offers (in percentage terms). Given the small size of even the highest stake (i.e., \$15), not surprisingly, the author concluded that with small stakes, fairness considerations dominate monetary considerations.

List and Cherry (2000) used a multiple-round design similar to that of Slonim and Roth (1998), to analyse how ultimatum proposals are influenced by changes in stake size. The authors wanted to reduce subjects' tendency towards making symmetric offers. To do that, they developed an experimental design which lead subjects to form entitlements/obligations: Subjects had 45 min to answer 17 questions from a GMAT exam. If the proposers cannot answer more than 10 questions correctly, the stake will be \$20; if they can, it will be \$400. Later, subjects were divided into two roles as usual, i.e. proposer and responder; and each subject played the ultimatum game 10 times. Proposers played with a different (and randomly selected) responder each period.

In addition to Slonim and Roth's (1998) findings, List and Cherry (2000) found that learning occurs across rounds only if stake size is sufficiently large (see Fig. 1). This result is in line with Smith–Walker hypothesis in that subjects were involved in cognitively costly activities only when stakes are high. According to their results, rejections were less frequent in later periods. Moreover, there was no difference in behaviour across games with high and low stakes, as far as the first round behaviour is concerned.

In yet another study where a considerable variation in stake size was implemented, Munier and Zaharia (2003) investigated the effect of a fifty-fold increase in stake size on subject behaviour. The experiment was conducted in France and Romania, in 1996 and 1997, with the paper-and-pencil method. Since the experiment was conducted in two different countries using different currencies, a common currency, ECU (experimental currency unit) was used; and the exchange rate was set as 1000 ECU = 40 French Francs = 40,000 Lei. The experiment was carried out with 1000 ECU for 5 periods and with 50,000 ECU for another 5 periods. At the time of the experiment, 50,000

ECU was equal to four times average monthly income of the participants in Romania. Proposers wrote their offers on a paper called the *offer bulletin* whereas responders wrote their answers on another paper called the *decision bulletin*; and both bulletins were given to the experimenter. These offers and answers were paired up randomly to determine subjects' earnings. Throughout the experiment, the role of each subject was fixed. This design choice enabled the authors to test for possible learning effects. To make sure that amounts are thoroughly considered each time, subjects were asked to write their offers/responses in ECU currency and their own currency on another paper. In this fashion, subjects converted ECU to their own currency each time they make a decision. Therefore, they would be aware of how much money their decision involved in their local currency. Results showed that, in both countries, the average minimum acceptable offer was significantly lower when the stake was 50,000 ECU. Again, in both countries, proposers offered more and converged to a 50–50% allocation when the stake was 50,000 ECU.

In more than two decades after the first ultimatum game experiment, dozens of studies were conducted. Hence, there was enough data for a meta-analysis. Meta-analysis studies have greater statistical power (to detect an effect) than individual studies and are less prone to biases than individual studies. Oosterbeek et al. (2004) carried out a meta-analysis on 75 experiments from 37 papers published from 1982 to 2003. Stake size was one of the variables they investigated. Their study showed that (i) as stake size increases, the ultimatum offers decrease, (ii) there is no learning effect in repeated ultimatum games, and (iii) ultimatum offers are higher when strategy method¹⁵ is used. From the responder perspective, they found that rejections decrease as stake size increases. Clear results they found on the effect of stake-size is encouraging for this line of research.

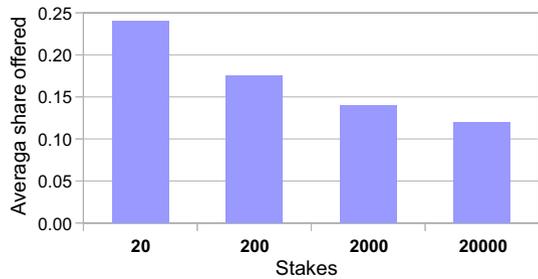
Carpenter et al. (2005), in resemblance with Forsythe et al. (1994), analysed the effects of stake size on subject behaviour in the ultimatum game by varying the stake between \$10 and \$100. As an advantage over many other studies, many individual and demographic variables were also controlled.¹⁶ In only one of the 39 ultimatum games, an offer was rejected. The average offer in \$10-game was 45% and the average offer in \$100-game was 41%. Median offers were 50 and 45%, respectively. Consequently, the study reached the conclusion that in ultimatum games, stake-size does not influence proposer behaviour. In light of earlier research which found that the variation used here is not sufficiently salient and very asymmetric offers are rare in standard ultimatum game experiments, this finding is expected.

As of 2007, almost all studies testing the stake-size effect in ultimatum game experiments to date have used student subjects. In one of the first attempts to change this to reach more external validity, Fu et al. (2007) used a nationally representative subject pool and controlled many socio-economic variables. One of the issues they focus on

¹⁵ In the strategy method, subjects declare their decisions for all their possible roles and for all possible situations beforehand. For example, in an ultimatum game, they announce what will be their offer when they are proposer, what will be their answer to all possible offers. After that, roles of players are determined, they are paired up with each other and payoffs are determined according to their strategies.

¹⁶ A recent study by Griffin et al. (2012) also investigated racial differences in ultimatum game behaviour. The authors varied the stake-size between \$5 and \$100. They concluded that stake-size did not significantly affect responders' acceptance decisions.

Fig. 2 Andersen et al. (2011)—offer proportions across stakes. This figure is regenerated from Fig. 1 in Andersen et al. (2011)



is the effect of monetary stakes on ultimatum game behaviour. They had around 800 adults (mean age, 41 years), randomly selected from the Taiwanese population to play the ultimatum game. Half of these participants played a low-stake game (NT \$200) and the other half played a high-stake game (NT \$1000). The researchers reported that at the time of the study, the ongoing hourly wage rate in Taiwan was NT \$100 and the survey took about 30 min to complete. They reported that raising the stakes reduces the frequency of high (fair) offers and the frequency of rejection of non-negligible offers, which are both in line with Smith–Walker hypothesis.

Most of the studies investigating stake-size effects in ultimatum games focused on proposer behaviour. On the other hand, Andersen et al. (2011) focused on how responders would behave (i.e., respond to low offers) when the stake size is *considerably* large. This question was not given a satisfactory answer before since proposers in ultimatum games almost never make very low offers. In the current study, even if the authors used very high stakes they must have still believed that they may not observe very low offers; and this is probably why they explained the reasoning behind the subgame perfect equilibrium behaviour (i.e. offering almost zero) in the instructions they gave to proposers.¹⁷

The experiment was conducted in eight villages located in Northwest India with 458 subjects. Each subject played the game once with a randomly assigned role (proposer or responder). Stakes were Rp 20 (roughly 1.6 h salary), Rp 200, Rp 2000, and Rp 20,000. As it can easily be calculated, the highest stake, Rp 20,000, is equivalent to a 1600 h salary and this was the largest stake (in real terms) used in ultimatum games so far.

Their results showed that ultimatum offers decrease as stake size increases (see Fig. 2). 88 % of the offers were below 30 % of the stake size.

More precisely, 20 % of the offers were below 10 % of the stake size; 28 % of the offers were between 10 and 20 %, and 40 % of the offers were between 20 and 30 %. Average offers were higher (in % terms) when the stake size was Rp 20. Another important result is that the rejection rate fell—for proportionally identical values—with larger stake sizes (see Table 3).

¹⁷ Since the aim of the study is analysing the behaviour of the responders, the possible experimenter demand effects resulting from the mentioned explanation does not pose a problem.

Table 3 Andersen et al. (2011)—rejection rates across stakes

	20 Rps	200 Rps	2000 Rps	20,000 Rps
% of rejections	46.43	36.00	19.57	0.00

This table is regenerated from Table 2 in Andersen et al. (2011) focusing on the “no wealth” treatment

Among all ultimatum game experiments in the literature, proposer and responder behaviour observed in the Rp 20,000 game in this study is the one closest to subgame perfect Nash equilibrium predictions. This result is important since it implies that predictions of theoretical models may hold true if sufficiently high amounts are at stake, which is in line with hypotheses in Rabin (1993), Smith and Walker (1993), and Bolton and Ockenfels (2000). We think that running more experiments with liberally high (and real) stakes such as the one in Andersen et al. (2011) is important to reach more informative results on this important issue.

Another recent study on the effect of stake size on subject behaviour was in stark contrast with Andersen et al. (2011) in terms of their stake-size choice. Amir et al. (2012) ran online ultimatum, dictator, trust, and public good games on Amazon’s Mechanical Turk (Mturk) platform with 756 subjects. Treatments with real incentives used \$1 as the stake size. Their results showed that proposers offered on average 49.7% of the stake and responders specified their minimum acceptable offer to be on average 33.7% of the stake. Amir et al. (2012) concluded that there was no significant difference between their results and the results of laboratory experiments using higher stake sizes. This result was valid for all four games they considered. It is important for experimental economists since it may have implications for experimenters’ choices of incentive levels: if it is true that using \$50 and \$5 will not make a difference, why would one use \$50 after all? Nevertheless, we think that these results should be interpreted with caution since Mturk is an online platform which means the opportunity cost for participating in this experiment was probably significantly lower than participating in a laboratory experiment and Mturk participants are probably more used to small-stake decisions than laboratory participants are.

The last study we want to mention in this subsection differs from the earlier ones in one important aspect. Zhou et al. (2014) studied the (ab/pre)sence of stake-size effects using fMRI data from a series of ultimatum games. Their study can be thought as a (neuro)experimental test of the prediction “as the stake size increases monetary considerations dominate fairness considerations” (see Rabin 1993; Bolton 1991). Experimental sessions were conducted in China and 28 people participated in the study. The low stake size was between 4 and 6 Yen and the high stake was between 400 and 600 Yen. At the behavioural level, the authors found a support for the hypothesis mentioned above: rejection rates for (comparable) unfair offers were significantly lower under high stakes than in low stakes. Another interesting finding was related to the identity of the proposers. The authors had computers and real human beings as proposers in different treatments. They reported that the stake-size effect on the rejection behaviour of responders mentioned above was valid only in the “human” condition. This implies that “. . . the modulation effect of fairness by monetary incentives only exists in social interactions situations”.

3.2 Other Bargaining Games

Although the ultimatum game is the most popular bargaining protocol in experiments, there are some studies using other bargaining protocols such as Nash bargaining, alternating offers bargaining, and unstructured bargaining. Here, we focus on studies where stake-size is varied.

Fouraker and Siegel (1963) in what can be considered as one of the first experimental economics studies conducted experiments on buyer-seller negotiations frequently observed in real life. In these *market games*, the buyer sets the price and the seller chooses the amount of goods he wants to sell at that price. Subjects played this game repetitively. Until the last period, low stakes were used and in the last period high stakes were introduced. Fouraker and Siegel's analyses showed that the average prices set by buyers and the average amounts set by sellers do not differ across games with lower and higher stakes. However, the variance of these variables was four times higher in the games with lower stakes, which is in line with the Smith–Walker hypothesis.¹⁸

A decade after Ståhl (1972) wrote his piece on sequential bargaining, Rubinstein (1982) published his seminal article on alternating offers bargaining. This paper had a huge impact on bargaining research and soon the Rubinstein–Ståhl (alternating offers) bargaining model became the most popular non-cooperative model of bargaining. In the two-player Rubinstein–Ståhl bargaining model, one of the players makes the first offer. If the other player accepts the offer, the game ends; if he does not, then he makes an offer to the other player. The game continues with players taking turns in making offers (in an alternating fashion) until an agreement is reached. One theoretical prediction is that the players will agree in the first period (known as the *immediate agreement result*). The distribution of the pie in this agreement is also determined by this prediction.

A natural question for many academics was the predictive power of Rubinstein–Ståhl bargaining model. In their paper where they tested the validity of subgame perfect Nash equilibrium prediction in the Rubinstein–Ståhl bargaining model with a two-period experimental design, Binmore et al. (1985) discovered that most subjects behaved in line with the equilibrium prediction.

Soon after Binmore et al. (1985), some researchers rightfully thought that an experiment using a two-period game is not good enough to test the predictive power of alternating offers bargaining model. Neelin et al. (1988), devised two, three, and five period designs. They also varied the pie-size to measure the robustness of their results to variations in the pie-size. The part of this experiment that is related to our study is the one where thirty undergraduate students played the Rubinstein–Ståhl bargaining game with \$5 and \$15 pies. The results showed that average offers (34 and 34 %) and average rejection rates (26 and 30 %) did not differ across games with different pie sizes.

In a study similar to Binmore et al. (1985), Güth and Tietz (1987) investigated subjects' behaviour in a two-period game in which if no agreement is reached in the

¹⁸ It should be noted that the actual aim of this experiment was not to measure the effects of the amount of money used on the subjects' behaviour. Accordingly, the change in stakes was implemented only in the last period. Factors such as experience, last period, and boredom effects might have introduced confounds.

first period, the value of the pie falls to 0.1 or 0.9 of its original value in the second period. In this experiment, each player played both roles (first making an offer, then receiving an offer) once. Researchers varied the pie (e.g., DM5, DM15, and DM35) to measure the effect of pie-size on behaviour. Increasing it from DM5 to DM35 had different effects on behaviour in games with a 0.1 discount factor than in games with a 0.9 discount factor. However, all in all, varying the pie-size did not result in any substantial change in behaviour. This study is important in that its results raised an interesting question about a possible interaction between time preferences and stake-size.

[Sonnegård \(1996\)](#) studied the effects of (i) assignment methods (i.e., random selection, contest, etc.) that determine the first proposer in different ways, (ii) labels and frames in instructions, and (iii) stake-size in bargaining games. He used a two-period bargaining game with alternating offers. In case of rejection of the first offer, a considerable portion (80%) of the pie is lost before the second period. The part of the experiment we are interested in here is the one where the author used 50 Swedish Crowns (equivalent of \$6 on the date of the experiment) as the low stake and 500 Swedish Crowns (equivalent of \$60) as the high stake. His results showed that fewer disagreements were observed when the stake was higher.

Up to now, all studies we mentioned are either controlled lab (or field) experiments or meta-analysis studies that used controlled experiments. In that aspect, [Goldreich and Pomorski's \(2011\)](#) study is a rare and precious one in the bargaining literature: they used a data set acquired from online poker tournaments and investigated *which type of players* initiated the negotiation process more frequently and *under what circumstances* were the negotiations successful. These poker tournaments start with 900 players and as players who run out of chips quit, the number of players are reduced to nine (or less) towards the end of the game. When there are nine or less players remain, any one of the players can propose a distribution of the pot as an alternative to resuming the game; and the players can negotiate on the basis of this offer. The bargaining protocol is free-form.

[Goldreich and Pomorski \(2011\)](#) collected data between April and July in 2007 and created a data set comprising of 1246 tournaments. The average prize pool in these tournaments was around \$80,000 and the maximum prize was over \$1,000,000. Therefore, in terms of stake-sizes it covers this study is a very rare one. In this data set, proposals were made and a consecutive agreement was reached in 31% of the tournaments and proposals were made but no agreement was reached in 34% of the tournaments.

The results of regression analysis conducted on the effects of the prize pool (stake size) on the probability of a bargaining proposal or the likelihood of an agreement are surprising. In five regressions using different combinations of variables the size of the prize pool had a quite small yet statistically significant effect on the aforementioned dependent variables. However, when the researchers analysed a sub-data-set that comprised of the tournaments with prize pools less than \$100,000, they discovered that the size of the prize pool did not have any significant effect on the dependent variables. This is a very interesting result given the extraordinary large stakes involved in these tournaments: it is in line with the results of many lab experiments that vary stake-size

mildly (e.g., \$5 vs. \$10, \$10 vs. \$50 etc.) and in contrast with others that introduce very large stakes (e.g., Andersen et al. 2011).

Güth and Kirchkamp (2012) studied the effect of stake-size in the *Yes–No game*. The Yes–No game is a variant of the ultimatum game where the responder does not know the offer when responding. In Straub and Murnighan (1995), offers were known by responders but stakes were not; and in Güth and Kirchkamp (2012), stakes were known by responders but offers not.¹⁹ Two different stake sizes were €100 and €1000. In addition, the authors also incorporated data from Gehrig et al. (2007) where €20 was used. Another important difference from many other studies is that, Güth and Kirchkamp used a within subject design to investigate stake-size effects, i.e. each player made a decision both for a €100 and a €1000 stake. In the pooled data, they found that the average offered share was 0.376 of the stake in the low-stake (€100) game whereas it was 0.364 of the stake in the high-stake (€1000) game. The difference is very small but highly significant. On the other hand, the average rejection rate was 6.6% in the low-stake game whereas it was 4.5% in the high-stake game. The difference is again small but yet again significant. As the authors acknowledged (see Güth and Kirchkamp 2012, p. 664), the statistical significances of these economically small effects may be due to the large number (1175) of subjects in their study.

Karagözoğlu and Riedl (2015) investigated how subjects' fairness judgements were formed and later influenced bargaining under different circumstances. They employed a bargaining problem with joint production.²⁰ Accordingly, pairs were given a real-effort task and their joint performance influenced the amount which they later will be bargaining over. The main factors that separate Karagözoğlu and Riedl (2015) from most others in the literature is that this study used unstructured bargaining protocol and the bargaining pie was not exogenously given but rather produced by subjects. In unstructured bargaining, subjects in each pair are given a fix period of time to reach an agreement. If they can agree within this time period, they get their agreed shares of the pie. If they do not, parties receive their disagreement payoffs.

In Karagözoğlu and Riedl (2015) study, stakes were determined by the joint performance of subjects in every group and could take three values: €9, €13.5 or €18. The authors checked the robustness of their results to changes in stake-size. Their results show that average fairness judgements, average time needed to reach an agreement, and average agreed shares remain the same under different stakes. Nevertheless, these results should be interpreted with caution since the high stake was only twice the low stake.

Camerer et al. (2015) studied dynamic unstructured bargaining with deadlines and one-sided private information. The major novelty of their study is that using Myerson's (1979, 1984) *revelation principle*, they derived some theoretical predictions. One prediction that is of interest for us stipulates that the disagreement rates should be monotonically decreasing in pie size. In the experiment, there are six possible pie

¹⁹ The authors used different media (e.g., lab, newspaper, internet) and reached different subject pools (e.g., newspaper readers, business executives, students). They also used different payment methods (e.g., paying everyone their full earnings, randomized payment).

²⁰ For a comprehensive survey of experiments on bargaining games with joint production, see Karagözoğlu (2012).

sizes ranging from \$1 to \$6. Pie size is known to one of the players whereas the other player knows that it is drawn randomly from a commonly known discrete uniform distribution. Randomly matched informed and uninformed players bargain for 10 s in an unstructured fashion. This is repeated for 120 periods with stranger matching protocol. Subjects' payoffs were calculated on the basis of their profits in randomly chosen 15 % of the periods plus a \$5 show-up fee. There were 110 subjects from Caltech and UCLA. Results indicate that disagreement rates indeed monotonically decrease in pie size. Moreover, they found a stake-size effect on the distribution of the pie: when the pie-size is small or medium, equal-splits were more frequently observed; and when it is high, informed parties received more than half of the pie. It is very striking (and in stark contrast with all of the earlier studies) that the authors obtained stake-size effects with such a small variation in stake size. Their findings may be considered as a support for the argument in [Camerer and Hogarth \(1999\)](#), which stipulated that stake-size effects may vary across different tasks/games.

4 Distribution Games

In this section we review experimental studies on the dictator game, the trust game and variants, the public good game, and some other games where authors varied the size of the stakes. These games involve a certain distribution problem yet they are not bargaining problems. Hence, we categorize them as distribution games.²¹

4.1 Dictator Games

In the standard dictator game, the decision of how to divide the money belongs only to the proposer, i.e., all initiative is on his hands. The proposer (or dictator) determines the division of the pie and it is implemented. The other player has no chance to make an alternative offer or to reject the offer at hand. It is expected that completely selfish players will take all the money for themselves. In contrast with ultimatum games, there is no fear of rejection for the proposer in dictator games. Therefore, if and when a proposer shares the pie with the other player in the dictator game, this is interpreted as an indication of fair-mindedness ([Camerer 2003](#), pp. 43–100).²² The results of many dictator game experiments show that proposers (or dictators) usually offer non-negligible shares (about 20–30 %) to other players.

[Sefton \(1992\)](#) is one of the earliest studies focusing on the implications of using “random pay” method in economic experiments. This method proposes paying only a certain percentage of subjects as opposed to full pay (i.e., paying all subjects). In testing whether random pay leads to the same behavioural patterns as with full pay, [Sefton \(1992\)](#) used the dictator game as a test bed. In the random pay treatment, only 25 % of subjects were paid their shares from a dictator game that involves \$5, whereas

²¹ “Non-bargaining distribution games” might have been more informative. We are aware of the fact that these games also involve an efficiency component. At the cost of using the label “distribution games” in a way possibly different than the standard one, we use it to differentiate them from bargaining games.

²² For alternative/opposing arguments, the reader is referred to [Bardsley \(2008\)](#) and [List \(2007\)](#).

in the full pay treatment all subjects were paid. Results clearly rejected the equality of shares given by the dictators across two treatments. In particular, subjects were more generous in the random pay treatment.

Cherry et al. (2002), in their frequently cited article, made two alterations to the well-known dictator game and analysed the effect of these alterations on behaviour. In the standard dictator game, the dictator role is assigned to one of the players in a random fashion. Another important point is that the standard version is not double-blind, i.e. it allows researchers to identify subjects with their decisions. Cherry et al. (2002) analysed the possible effects of changing these two design details on subjects' behaviour. They assigned the dictator role on the basis of subjects' performances in a contest and used a (double-blind) design that did not allow experimenters to identify subjects' decisions. Stake size was determined at the contest stage (answering trivia type of questions). In this stage, subjects were not informed about the dictator game they will play in the second stage. In the dictator game stage, stake size was \$40 if the more successful subject in the contest stage has 10 or more correct answers out of 17 and it was \$10 if he/she had less than 10 correct answers.

The main results of this study were as follows: (i) in the control treatment in which dictatorship was assigned randomly, 19 % of dictators did not give any share to the other player in the \$10-game; this rate was 15 % in the \$40-game, (ii) in the treatment in which dictators were determined according to their performances in the contest, 79 % of all dictators did not give any share to the other player in the \$10-game; this rate was 70 % in the \$40-game, and (iii) in the double-blind treatment in which dictators were, again, determined according to their performance in the contest, 97 % of all dictators did not give any share to the other player, when the stake size was \$40. Cumulative distributions of offers in different treatments also looked identical for \$10 game and \$40 game. These results are in line with a hypothesis that the behaviour in the dictator game is not systematically influenced by changes in stake size, even when dictator roles are *earned*. The only difference related to stake size was that the effect of the double-blind protocol (e.g., offers converging to zero) when the stake was earned was less pronounced in the low-stake compared to the high-stake game.

One of the important questions investigated in Carpenter et al. (2005) is the effect of stake size on behaviour in a dictator game. These authors conducted dictator game experiments with stake sizes, \$10 and \$100. The experimental design is similar to the one in Forsythe et al. (1994). Besides the use of a \$100 large stake, another novelty of this experiment is that the authors controlled many individual-level variables such as Machiavellian tendencies, risk aversion etc. and many demographic characteristics such as age, sex, education level, income level of family, ethnic origin, number of siblings etc. Since \$100 is a noteworthy amount, we suppose this may have limited the number of subjects: 19 players played the game with a \$10 stake and 21 players with a \$100 stake. In the game with a \$10 stake, the average share given to second players was higher (33 %) than in the game with a \$100 stake (25 %). Nevertheless, this result was due to an outlier in the former game, which gave all of \$10 to the other player. After removing this outlier from the data set, all statistical tests lead to the same result: there is no effect of stake size on dictator behaviour. On the other hand, the authors conducted the same analysis using median instead of mean: median share given in the \$10-game was 40 % while it was 20 % in the \$100-game. However, the authors

do not report any statistical significance about this difference. On the other hand, something worth mentioning is that authors found that some individual characteristics such as race, number of siblings, risk attitudes, family income, endowments etc. have a potential to explain behaviour.

List and Cherry (2008) can be seen as a variant of Cherry et al. (2002). These researchers used the design in Hoffman et al. (1994) as a control treatment and carried out two more treatments. In the *asymmetric treatment*, subjects chosen as dictators worked on a 45 min test to generate the stake that is going to be divided afterwards. In the *symmetric treatment*, both players worked on the test to generate the stake. Ten or more correct answers in the test generated \$100 and less than ten correct answers generated \$20. Even though the data is consistent with the hypothesis that an increase in stake size increases shares given by dictators (at a rate less than the rate of increase), statistical tests do not offer any such evidence since the variation in data set is so small.

Ben-Ner et al. (2008) studied the giving behaviour in dictator games with hypothetical versus real money (\$10). Their most important contribution, in our opinion, lies in the potentially relevant variables they controlled. Using questionnaires, they controlled for personality traits (big five) and cognitive abilities. Their findings suggest that the giving behaviour of an average subject was not different across real versus hypothetical conditions. Nevertheless, they observed that two factors, *agreeableness* and *extraversion*, interacted with the payment condition (hypothetical vs. real) significantly. More precisely, subjects with high agreeableness scores gave more when playing with real money, whereas subjects with high extraversion scores gave more when playing with hypothetical money. These valuable observations would not be obtained if these factors were not elicited, which also highlights the importance of the subject recruitment and data collection method we discuss in this paper.

More recently, Engel (2011) conducted a meta-analysis with a rich data set from more than a hundred dictator game experiments. In the data the researcher used to investigate stake-size effects, the stake varied from \$0 to \$130, with a mean of \$21.77. Meta-analysis results using the whole data set found an insignificant stake-size effect. Yet another surprising result is that hypothetical versus real stakes were reported to lead to similar behaviour. Nevertheless, when the analysis is restricted to those treatments that manipulated stake-size, then there is a significant stake-size effect: higher stakes reduced dictators' willingness to give.

A study by Bellemare et al. (2013) focused on the sensitivity of guilt aversion to changes in the stake size in dictator game experiments. In their *MAIN* treatment, they had 142 dictators. Each dictator was confronted with three games, which differed in stake size. The details are shown in Table 4.

It is clear that if the dictator cares only about his own monetary payoff, he'd choose r in all three games. If the dictator has guilt aversion, then he is expected to choose l . Finally, the dictator may be guilt averse but the degree of his guilt aversion may be sensitive to stake. The authors elicit passive players' first-order beliefs about how many dictators out of ten would choose option l . Dictators make their decisions for each game knowing that they will be randomly matched to the first-order belief of one passive player for each decision they made.

The authors reported that 101 of the dictators are stake-sensitive in their guilt aversion. And most subjects became more selfish (or less guilt averse) as stake-size

Table 4 Bellemare et al. (2013)—experimental details

Games/decisions	Option <i>l</i>	Option <i>r</i>
Game 1	50 DK for the dictator, 48 DK for the passive player	54 DK for the dictator, 22 DK for the passive player
Game 2	100 DK for the dictator, 96 DK for the passive player	108 DK for the dictator, 44 DK for the passive player
Game 3	200 DK for the dictator, 192 DK for the passive player	216 DK for the dictator, 88 DK for the passive player

DK Danish Krone

increased. Only 39 of the dictators were stake-insensitive. Moreover, their structural estimation results confirmed the same result: people become more selfish as the stakes increased. It is remarkable that with a stake-size variation not larger than many earlier studies, Bellemare et al. found clear results stake-size effects.

Bardsley (2008) and List (2007) reported that a simple manipulation of dictators' strategy spaces (i.e., allowing them to take away from the other agent), the giving behaviour in the dictator game can be altered (e.g., dictators in the new game become less generous). Along similar lines Leibbrandt et al. (2015) manipulated whether dictators could GIVE to or TAKE from another individual and whether the endowment was moderate (LOW) or very large (HIGH). Their experiment was conducted in rural Bangladesh. Low stake was roughly a day's earning (100 tk), whereas the high stake was equal to a five months' earning (10,000 tk). The authors observed that absolute money amounts offered/left by dictators are higher under high stake than in low stake (1816 tk vs. 27 tk). However, the proportion of the stake offered/left by dictators was lower under high-stake than in low stake (18 vs. 27 %, $p = 0.087$). Understandably, the proportion of dictators who chose equal-split was also affected by the stake-size manipulation. This ratio is 9.5 % for the high stake, whereas it is 21 % for the low stake. More interestingly, the authors observed an interaction effect between treatments. In particular, stake-size manipulation influenced dictator behaviour in the GIVE frame but not in the TAKE frame. On average, dictators allocated 25 % in low stake GIVE condition, but only 3.7 % in high stake GIVE condition ($p < 0.001$, MW test). In contrast, these average figures were not statistically different across low-stake TAKE and high-stake TAKE conditions.

4.2 Trust Games and Variants

In the trust game (Berg et al. 1995), the first player (sender) chooses to send some share—giving nothing is also allowed—of a stake to the other player (receiver). Then, this amount is multiplied with a number larger than one—in this case, this number is three—and given to the receiver. After that, the receiver may send back some part (again, giving nothing is allowed) of this larger amount to the sender and the game ends.

Parco et al. (2002) analysed the effect of financial incentives on decisions in the centipede game. The *centipede game* is a finite sequential game where two players,

in an alternating fashion, decide on whether to take a slightly larger share of a slowly increasing pot, or to pass the pot to the other player. Payoffs in the game are such that if a player passes the pot to his opponent and the opponent takes the pot in the next round, the player who passed the pot receives slightly less compared to the amount he would have received if he had taken the pot. In [Parco et al. \(2002\)](#), a three-player centipede game with 60 turns was played by two groups of 15 subjects. Each group was divided into 5 sub-groups with 3 players each. Results showed that, compared to the game with a lower stake, the likelihood of the game ending in the first period (i.e., the first player grabbing most of the pie in his first turn) was fifteen times higher with a higher stake. So much so that, in the game with a small stake size, only 25.2% of the first players decided to end the game, while this rate was 83.4% in the case with a higher stake size.

[Ho and Weigelt \(2005\)](#) conducted multi-stage trust game experiments in United States, China and Singapore with 386 subjects. One difference from other trust game experiments here is that they used the strategy method and hence collected data on subjects' strategies (rather than actions). Subjects in their experiment played the multi-stage trust game for 10 rounds. The high stake was 10 times the low stake. They found that subjects exhibited significantly more trusting behaviour in the high-stake game. Higher stake was also associated with more trustworthiness. The authors conducted only statistical tests whereas some regression analyses might have also been useful since there are many other factors that might be influencing subject behaviour some of which were controlled.

In their experiment conducted in Bangladesh, [Johansson-Stenman et al. \(2005\)](#) analysed the effect of stake-size on subject behaviour in the *trust game*. They divided subjects into three groups of 64, 59, and 62 and in each group they gave the sender 40 Taka, 200 Taka, and 1000 Taka, respectively. Hence, there were three different stake sizes. Their results showed that amounts first players sent fell with increasing stakes whereas amounts second players sent back were not affected by that increase. This is similar to some of the results these authors found in a more recent study ([Johansson-Stenman et al. 2013](#)) they conducted in rural Bangladesh, that is people trust others less when stakes are high. It is interesting to see that the change in stake-size has differing effects on sender and receiver behaviour. We believe that these results call for further replication studies with more controls to understand the effect of stake-size on behaviour.

As of 2008, there were plenty of studies focusing on trust behaviour. Some only used self-reported measures of trust from surveys and some used the trust game paradigm in incentivized experiments. For instance [Glaeser et al. \(2000\)](#), in a seminal article, reported that the correlation between self-reported measures of trust and behaviour in the trust game is rather weak. In contrast, [Buchan and Croson \(2004\)](#) reported that subject behaviour in a hypothetical (unpaid) trust game conducted in United States and China is very close to the behaviour observed in trust game experiments with real money at stake. Other authors using very small stake sizes (€2) along with slightly higher values (€8) found *no-difference* results (see [Sutter and Kocher 2007](#)). However, there was no study which implemented a hypothetical trust game and a trust game with real incentives together. A part of [Holm and Nystedt \(2008\)](#) study was dedicated to

this. In the HYPOTHETICAL treatment, there are no financial incentives whereas in the REAL treatment the initial stake in the trust game is 50 SEK (about \$7 at the time). The authors had further treatments (FEEDBACK vs. NO FEEDBACK) to disentangle some potential confounding factors. Nevertheless, the behavioural shifts due to the presence of real stakes were identical in these treatments. Real stakes did not influence receivers' behaviour but in contrast with the common belief that senders would be less trusting in games involving real stakes, senders were significantly more trusting. As the authors put it, the asymmetry present in the effect of real stakes may be in line with the Smith–Walker hypothesis: it's the senders who would increase their cognitive effort when real incentives are given; receivers are not involved in much of sophisticated cognitive activity anyway.

Similar to [Holm and Nystedt \(2008\)](#), [Fetchenhauer and Dunning \(2009\)](#) was also interested in the effects of real stakes in trust games. In one part of their paper (Study 2), they had hypothetical trust game (no real payment) and potentially real trust game (with a certain probability, subjects are paid based on their trust game decisions). These treatments were run in the Netherlands and 118 university students participated. In both treatments, subjects first earned €7.50 by performing a certain simple task; and then they were asked to take decisions some of which concern this money they just earned. In particular, they were asked to decide (both as a sender and a receiver) in a modified version of the trust game where sender's decision is whether to send the €7.50 or not, and the receiver's decision is to keep all of €30 (if €7.50 is sent, the receiver receives €30) or split it equally between himself and the sender. Subjects were also asked to guess whether the senders would return half of €30 or not. These guesses did not differ significantly between real and hypothetical treatments. However, similar to the corresponding finding in [Holm and Nystedt \(2008\)](#), senders were significantly more trusting with (potentially) real stakes than with hypothetical stakes: 37.7% (56.3%) sent the money in the hypothetical (real) condition.

[Naef and Schupp \(2009\)](#) came up with a new survey that aims to measure trust and connect participants' answers to survey questions to their behaviour in the trust game, using a large, representative participant pool from Germany.²³ Among other things, they also investigated the robustness of their results with respect to stake-size. Two possible stake sizes were €10 and €100. They found that the average transfer was €49.9 in the €100-game and €5.16 in the €10-game: the same ten-to-one ratio that they have in stake sizes. The distributions of transfers were not different either. Finally, the level of trustworthiness was, again, identical across two treatments. These results are very similar to the ones found in ultimatum or dictator game experiments using similar stake size variations.

[Johnson and Mislin \(2011\)](#) conducted a meta-analysis on the data from 162 trust game replications. More than 23,000 people from a variety of geographical locations participated in these studies. The authors investigated the effects of many variables

²³ There is a lot of important methodological details in [Naef and Schupp \(2009\)](#) study, which we cannot cover here, in this kind of a paper. Hence, the interested reader is referred to the original paper for such details.

such as payment method, experimental protocol, simulated versus real counter-part, geographical variation etc. They also controlled for the stake-size effect. On the point that is of primary interest for us, they found a clear result indicating that the size of the stake did not influence subject behaviour in any sense. Nevertheless, they note that in only eight of the studies their data set covers, the stake was higher than \$25 and only one in those eight, trust game earnings are guaranteed (i.e., not randomized). In light of our knowledge from the results on other experimental games, we can say that this level of variation might not have been sufficient for the presence of stake-size effects. It is fair to say that the question concerning the effect of stake-size on behaviour in trust games still needs further investigation.

In a very recent article, [Centorrino et al. \(2015\)](#) used a two-person trust game where senders observed short video clips of trustees before taking their decisions. Participating senders also rated these video clips along a number of dimensions. The most important dimension was the genuineness of the trustee's smile in the video clip. The authors varied the money at stake: the low stake was €4 and the high stake was €8. Their main finding is that trustee smiles in the high stake treatment are rated as more genuine and those trustees are perceived as being more trustworthy. Consequently, a higher percentage of senders sent money to these trustees. Nevertheless, the percentage of trustees returning any money to the senders is lower in the high stake game. These results are interesting in that such a small increase in the stake size appears to have generated sufficient incentives to create more genuine "acting" and more selfish behaviour.

4.3 Other Distribution Games

[Kocher et al. \(2008\)](#) conducted the first study on the effects of stakes on cooperation and punishment behaviour in a public good experiment. Public good game is a game played in groups. At the beginning, each subject in the group is given a certain amount of resource (i.e., private resources) and then they are simultaneously asked to decide how much of this resource they will keep for themselves and how much of the resources they will contribute to the production of a public good that would benefit everyone in the group. The sum of their contributions is multiplied by a number larger than one (so that their contributions at least collectively pays off). Under reasonable parameter values, the Nash equilibrium of this game is: none of the players contribute to the production of public good (i.e., the well-known free-rider problem). As for the version of public good game with the costly punishment option: after the contribution stage, every player has the option to punish the other players by paying a certain cost, which reduces their individual gains.

Both of these versions (with and without the punishment option) were played in South Africa by 120 high school students, in groups of three. Every subject played the game only once. The low stake used in the game was one fifth of the high stake. Neither the average contribution of the players to the public good production nor the variation in this amount was affected by the stake size. Additionally, the authors found out that the punishment behaviour of the players were not affected by the stakes either. The

results are clear and surprising. Nevertheless, the sample is somewhat non-standard: high school students in South Africa.²⁴ It would be interesting to see the results of replication studies in other geographical locations and with university students or adults.

Another study using data from a real-life event is [Oberholzer-Gee et al. \(2010\)](#). They investigated how certain variables affect behaviour of players by observing 630 players appearing in 315 games, in 105 episodes of a TV game show, ‘*Friend or Foe?*’, which was broadcasted for two seasons starting in 2002.²⁵ This game initially involves players contributing to a common prize pool by answering general knowledge questions and this stage is followed by players playing a strategic game to decide how to share the commonly generated prize. To be more precise, if both players select the ‘friend’ option then the prize is divided equally; if both players go with the ‘foe’ option neither of them gets the prize, and if one of them goes with ‘friend’ and the other with ‘foe’, the player who selects the ‘foe’ option gets all the prize and the ‘friend’ gets none. In accordance with their success in answering general knowledge questions, the prize pool ranges from \$200 to \$20,000. The analysis showed that among the 630 players, 45% of them went for the ‘friend’ strategy. More importantly, the percentage of players who opted for ‘friend’ option in the whole population of players was not affected by the amount of the money at stake.²⁶

[Fehr et al. \(2014\)](#) investigated if high monetary incentives eliminated reciprocal behaviour or not. For this purpose, they ran gift-exchange market and complete contract market experiments in Moscow. High monetary incentives used in this study corresponded to roughly 2–3 months’ salary (at the time of the experiment) in Russia. The low monetary incentive was one-tenth of the high monetary incentive. Games were repeated ten times to control for potential experience and learning effects. The results revealed that subject behaviour was not affected by the size of the stake. The agreed salary values were similar across high and low monetary incentives. Subjects who assumed the role of the employer—in both low and high incentives—offered the subjects who assumed the role of the employee roughly three times the salary that would emerge in a competitive market. The size of the stake did not appear to have an effect on the levels of efforts exerted either. Overall, the results of this study—despite the presence of a very high monetary incentive—were not in line with the hypothesis that higher stakes lead fair or reciprocal behaviour to disappear.

²⁴ For instance, we know that culture plays a role in shaping subjects’ behaviour in social interactions ([Henrich et al. 2001](#); [Herrmann et al. 2008](#)), which may make it difficult to generalize these findings. See also [Raihani et al. \(2013\)](#) for a recent study finding behavioural differences across subjects in USA and India in online dictator games with \$1, \$5, and \$10 stakes.

²⁵ This study and aforementioned [Goldreich and Pomorski \(2011\)](#) study are not experiments conducted in a controlled lab environment. Such experiments are called quasi-experiments in the literature.

²⁶ In a recent paper very similar to this one, [Van den Assem et al. \(2012\)](#) analysed the cooperative behaviour in the television show, *Split or Steal*. The stake-size varies between a few dollars and \$175,000, averaging over \$20,000. The authors found only limited support for the hypothesis that cooperation will decrease as the stakes get larger.

5 Discussion

It is clear that the results of studies investigating the influence of stake-size on subjects' behaviour in bargaining and distribution game experiments are far from conclusive and present a rather mixed picture.

Here we discuss a subject recruitment method that could potentially be useful in answering the stake-size effect question. With this method, researchers will be able to gather information that is relevant for research questions periodically or even repetitively (before and after the experiment). This is expected to help researchers acquire a large set of potentially important control variables that will be crucial in avoiding confounding effects. A suggestion along the same lines also appears recently in [Güth \(2012\)](#), who argued "... rather than distinguishing bargaining only by individuals and groups, the astonishing heterogeneity of lab-observed bargaining behavior suggests a better control and categorization of individual and possibly group types. ... To elicit the appropriate characteristics, one may rely, for instance, on panels for recruiting participants which may provide a lot of socio-demographic background information, ...". Findings on the relevance of two personality factors (agreeableness and extraversion) for giving behaviour in dictator games with hypothetical versus real money reported in [Ben-Ner et al. \(2008\)](#) also support our arguments here.

In this subject recruitment system, students who want to participate in experiments will initially answer some questionnaires to shed light on their behaviour in the experiments they are to participate. These questionnaires will acquire information on subjects regarding their risk-taking tendencies (e.g., [Holt and Laury 2002](#); [Dohmen et al. 2011](#)), their fairness concerns [e.g., justice sensitivity scale ([Schmitt 1996](#)), justice centrality scale ([Dalbert et al. 1987](#)), belief in a just world scale ([Rubin and Peplau 1975](#); [Dalbert et al. 1987](#); [Dalbert 1999](#))], loss aversion tendencies (e.g., a loss aversion measurement scale), some characteristics that may relate to tendencies like altruism, selfishness [e.g., Big five personality traits ([McCrae et al. 2005](#)), Machiavellian inclinations questionnaire ([Christie and Geis 1970](#))], cognitive abilities (cognitive ability/aptitude test), and factors that may relate to negotiation or distribution behaviour such as gender, age, job/market experience, area of education, number of siblings, financial status and time/opportunity costs of subjects. There are many experimental economics and psychology studies reporting how each and every one of these factors affect individual behaviour in strategic decision-making situations.²⁷ Moreover, these factors should be controlled for in an econometric analysis based on a theoretical model whose essential elements were outlined in Sect. 2.

The students who volunteer as subjects can complete these questionnaires whenever they want, but reasonably before participating in an experiment, anonymously and receiving low monetary incentives. Additionally, some of these questionnaires (e.g. risk-taking and loss-aversion tendencies, financial status etc.) will be filled out periodically and therefore the researcher will be able to observe the changes—if any—

²⁷ It is worthwhile emphasizing that we are advocating to control only those variables for which there exists some form of theory and/or wide-spread empirical evidence for the link between them and the investigated behaviour.

in these individual dimensions.²⁸ It is important that the information can be obtained periodically/repetitively since the subjects risk attitudes and financial status may not be stable.

Thanks to the suggested subject recruitment and data collection system, the following primary benefits will be acquired: (i) the researcher will possess a rich and dynamic data on the relevant individual characteristics of his/her subject pool, (ii) the researcher will be able to save time and resources during experimental sessions, which is expected to have positive effects in reducing *boredom effects* and cutting down costs, and most importantly (iii) since many theoretically relevant factors are controlled, the results of experiments investigating the effect of stake-size on behaviour in bargaining and distribution games will be more sound and offer richer insights for the presence or absence of stake-size effects.

Another benefit of this mechanism is related to a problem that is commonly referred to as the *experimenter demand effect*. This phenomenon can be defined as the design of the experiment (e.g. questions asked, instructions etc.) affecting the behaviour of the subject through its potentially hidden clues as to the experimental manipulation of the experiment and driving the subject not to act or behave as he/she normally would. This phenomenon appears when the subjects are asked extensive questions relating to the purpose of the experiment (e.g. asking many questions about fairness in a distribution experiment). The subject recruitment and data collection methods we propose are expected to reduce the risk of observing experimenter demand effect since they take the questions that may relate to the purpose of the experiment out of the experiment. Student volunteers will answer many control questionnaires that differ drastically in content and they will do this reasonably beforehand so that subjects are not likely to associate the purpose of the experiment with relevant questionnaires.

6 Conclusion

Experimental economics has been one of the fastest growing fields in economics in the last 30 years. We have examined the question of whether the size of monetary incentives has any effect on individual behaviour in experimental bargaining and distribution games.

Our observations can be summarized as follows: (i) possibly except for the ultimatum games, the number of the studies conducted is not sufficient to reach a definite conclusion on the effect of *stake size* on individual behaviour; (ii) the findings of experiments, including that of ultimatum experiments, are far from giving a clear idea about the effect of stake size on individual behaviour; (iii) uncontrolled individual characteristics such as co-existence of increasing relative risk aversion and increasing cost of fairness and small sample sizes could be responsible for these vague results; (iv) if there is indeed a stake-size effect, it can possibly be observed with the use of

²⁸ An important concern here would be anonymity of participants. Nevertheless, this would not be a major issue since (i) participants would be asked to generate participant id's that conceal their identity and (ii) payment methods that keep participant id's anonymous to the researcher already exist and are used.

very high stake sizes (e.g. [Andersen et al. 2011](#)); (v) the use of quasi-experiments and meta-analysis studies can be a promising venue.

In our opinion, these observations indicate that one of the things we need is more control in experiments. To reach clear results on the effect of *stake size* on individual behaviour, we need larger sample sizes, better subject recruitment systems, and data collection method that enables us to control potentially important variables of interest. The need for and the importance of control is better understood if we take into account the fact that such experiments are conducted in so many different geographical locations (e.g., rural India, Africa, Slovakia, France, Indonesia, Germany, Russia, United States etc.) and with very different subject pools (e.g., high-school students, college students, professional lawyers, unemployed low-income households etc.) because of the difficulties of conducting them with considerably high stakes in developed countries (see [Güth and Kirchkamp 2012](#), p. 660). All of these varying factors can also introduce confounds that may lead to misleading estimates for stake-size effects. We believe that the methods we have presented in this work have a potential to make a valuable contribution to this literature.

Another important factor, which makes it difficult to answer the “stake-size question” clearly is the variety of experimental designs and strategic interactions used in the studies reported. In addition to geographical locations and subject pools, information conditions, design details, important parameter values etc. vary across experiments. Some studies find stake size effects with variations at the magnitude of a few dollars, whereas some others cannot find despite a very large variation in the stake size. We believe that disentangling the types of strategic interaction and design elements that have an influence on the presence/absence of a stake-size effect is a challenging but very promising venue of research.

We also believe that an increase in the number of quasi-experiments using data from very high-stake events will be a good complement to lab experiments using the control strategies we mention above due to obvious pros and cons of these two methods discussed before. Finally, meta-analysis studies,²⁹ which promise results with greater precision and statistical power compared to individual experiments will be important tools for researchers.

Acknowledgments Emin Karagözoglu thanks The Scientific and Technological Research Council of Turkey (TÜBİTAK) for the research grant #111K499. The authors would like to thank two anonymous reviewers and an editor for their comments that improved the paper. The authors would also like to thank Gary Bolton and Urs Fischbacher for their comments and Mirza Trokic for proof-reading the manuscript.

²⁹ Recently, there has been an increased interest in meta-analysis studies in experimental economics. The reader is referred to [Cooper and Dutcher \(2011\)](#) for ultimatum games, [Engel \(2011\)](#) for dictator games, and [Johnson and Mislin \(2011\)](#) for trust games.

Appendix

See Tables 5 and 6.

Table 5 Studies reporting stake-size effects

Notes	Year	Reference	Game	Stake-size	Real/hypothetical	Subjects	Country
	1993	Nagel and Mitzkewitz	Ultimatum	DM1.20–DM7.20	Real	Students	Germany
	1995	Tompkinson and Bethwaite	Ultimatum	\$10 and \$10,000	Hypothetical	Students	New Zealand
	1995	Straub and Murningham	Ultimatum	\$10, \$30, \$50, \$60, \$80, \$100, \$1000, and \$1,000,000	Mixed	Professional Lawyers	Unites States/Canada
1	1998	Roth and Slonim	Ultimatum	60, 300, and 1500 Slovak Koruns	Real	Students	Slovakia
	2000	Dickinson	Ultimatum	\$1, \$2, \$4, \$7, \$15	Real	Students	United States
2	2000	List and Cherry	Ultimatum	\$20 versus \$400	Real	Students	United States
	2003	Munier and Zaharia	Ultimatum	40 versus 2000 French Franks	Real	Students	France and Romania
3	2004	Oosterbeek et al.	Ultimatum	Various	Mixed	Various	Various
	2007	Fu et al.	Ultimatum	NT\$200 versus NT\$1000	Real	Representative Population	Taiwan
4	2007	Gillis and Hettler	Ultimatum	\$10	Mixed	Students	United States
	2011	Andersen et al.	Ultimatum	Rp 20, Rp 200, Rp 2000, and Rp 20,000	Real	Rural Village Farmers	India
	2014	Zhou et al.	Ultimatum	4–6 Yuan (low) 400–600 Yuan (high)	Real	N/A	China
5	1963	Fouraker and Siegel	Bargaining	High versus low	Real	Students	United States
	1996	Sommegaard	Multi-stage bargaining	50 versus 500 Swedish Crowns	Real	Students	Sweden

Table 5 continued

Notes	Year	Reference	Game	Stake-size	Real/hypothetical	Subjects	Country
6	2012	Goldreich and Pomorski	Free-form bargaining	\$80,000–\$1,000,000	Real	Online poker players	International
	2012	Güth and Kirchkamp	Yes–No game	€100 versus €1000	Real	Newspaper readers, business executives, students	Germany
	2015	Camerer et al.	Unstructured bargaining	\$1–\$6	Real	Students	United States
	2002	Parco et al.	Centipede	\$0.01–\$25.60	Real	Students	United States
	2005	Ho and Weigelt	Trust	24 versus 320 Yuan	Real	Students	China
	2005	Johansson et al.	Trust	40, 200, and 1000 Taka	Real	Rural households	Bangladesh
	2008	Holm and Nystedt	Trust	50 Swedish Crowns	Mixed	Students	Sweden
	2009	Fetchenhauer and Dunning	Trust	€7.50	Mixed	Students	The Netherlands
	2015	Centorrino et al.	Trust	€4 versus €8	Real	Students	France
7	2012	Van den Assem et al.	Split-or steal	From a few \$s–\$175,000	Real	TV show participants	United States
	1992	Sefton	Dictator	\$5 (full pay versus random pay)	Real	Students	United States
	2011	Engel	Dictator	Various	Mixed	Various	International
	2013	Bellemare et al.	Dictator	98, 196, 392 Danish Kroners	Real	Students	Denmark
	2015	Leibbrandt et al.	Dictator	100 taka, 10,000 taka	Real	Rural Households	Bangladesh

1. Participants played the ultimatum game multiple rounds. The stake size effects were not present in all rounds
2. Participants played the ultimatum game multiple rounds. The stake size effects were not present in all rounds
3. This is a meta-analysis study
4. Differences in behaviour is not observed in the first round of play but observed in later rounds
5. The high stake was used only in the last round
6. This is a quasi-experimental study
7. This is a quasi-experimental study
8. This is a meta-analysis. Results are mixed. Some of them indicate that there is no significant stake-size effect

Table 6 Studies reporting no stake-size effects

Notes	Year	Reference	Game	Stake-size Variation	Real/hypothetical	Subjects	Country
1	1991	Roth et al.	Ultimatum	\$10 versus \$30	Real	Students	United States, Japan, Israel, Yugoslavia
	1994	Forsythe et al.	Ultimatum/dictator	\$5 versus \$10	Mixed	Students	United States
	1996	Hoffman et al.	Ultimatum	\$10 versus \$100	Real	Students	United States
	1999	Cameron	Ultimatum	Rp 5000, 40,000 and 200,000	Mixed	Students	Indonesia
	2005	Carpenter et al.	Ultimatum/dictator	\$10 versus \$100	Real	Students	United States
	2012	Amir et al.	Ultimatum/dictator/trust/public Good	\$0 versus \$1	Real	Various	Online
	1988	Neelin et al.	Alternating offers bargaining	\$5 versus \$15	Real	Students	United States
	1987	Güth and Tietz	Multi-stage bargaining	DM5, DM15, and DM35	Real	Students	Germany
	2015	Karagözoğlu and Riedl	Free-form bargaining	€9, €13.5, and €18	Real	Students	The Netherlands
	2002	Cherry et al.	Dictator	\$10 versus \$40	Real	Students	United States
	2008	Ben-Ner et al.	Dictator	\$10	Mixed	Students	United States
	2008	List and Cherry	Dictator	\$20 versus \$100	Real	Students	United States
	2007	Sutter and Kocher	Trust	€2 versus €8	Real	Various	Austria
	2009	Naef and Schupp	Trust	€10 versus €100	Real	Various	Germany

Table 6 continued

Notes	Year	Reference	Game	Stake-size variation	Real/hypothetical	Subjects	Country
2	2011	Johnson and Mislin	Trust	Various	Mixed	Various	Various
	2008	Kocher et al.	Public Goods	30 versus 150 S. African Rand	Real	High-school students	South Africa
3	2010	Oberholzer-Gee et al.	Friend or foe	\$200–\$20,000	Real	Various	United States
	2014	Fehr et al.	Gift-exchange/contract	High: 2–3 months income. Low: 1/10 of high	Real	Students	Russia

1. Stake variation was implemented only in the United States
2. This is a meta-analysis study
3. This is a quasi-experimental study

References

- Amir O, Rand DG, Gal YK (2012) Economic games on the internet: the effect of \$1 stakes. *Plos ONE* 7:e31461
- Andersen S, Harrison G, Lau M, Rutstrom E (2008) Risk aversion in game shows. In: Cox JC, Harrison GW (eds) *Risk aversion in experiments (Research in experimental economics, vol 12)*. Emerald Group Publishing, Bingley, pp 359–404
- Andersen S, Ertac S, Gneezy U, Hoffman M, List JA (2011) Stakes matter in ultimatum games. *Am Econ Rev* 101:3427–3439
- Baltussen G, Post T, van den Assem M (2008) Risky choice and the relative size of stakes. unpublished manuscript
- Bardsley N (2008) Dictator game giving: altruism or artefact? *Exp Econ* 11:1221–1233
- Bardsley N, Cubitt R, Loomes G, Moffat P, Starmer C, Sugden R (2010) *Experimental economics: rethinking the rules*. Princeton University Press, Princeton
- Bearden JN (2001) Ultimatum bargaining game experiments: the state of the art. Working paper
- Bellemare C, Sebald A, Suetens S (2013) Heterogeneous guilt aversion and incentive effects. Working paper
- Ben-Ner A, Kramer A, Levy O (2008) Economic and hypothetical dictator game experiments: incentive effects at the individual level. *J Socio-Econ* 37:1775–1784
- Berg J, Dickhout J, McKabe K (1995) Trust, reciprocity, and social-history. *Games Econ Behav* 10:122–142
- Binmore K, Shaked A, Sutton J (1985) Testing noncooperative bargaining theory: a preliminary study. *Am Econ Rev* 75:1178–1180
- Binswanger HP (1981) Attitudes toward risk: theoretical implications of an experiment in rural India. *Econ J* 91:867–890
- Bolle F (1990) High reward experiments without high expenditure for the experimenter? *J Econ Psychol* 11:157–167
- Bolton GE (1991) A comparative model of bargaining: theory and evidence. *Am Econ Rev* 81:1096–1136
- Bolton GE, Ockenfels A (2000) ERC: a theory of equity, reciprocity, and competition. *Am Econ Rev* 90:166–193
- Bolton GE, DeBruyn A (2008) Estimating the influence of fairness on bargaining behavior. *Manag Sci* 54:1774–1797
- Bombardini M, Trebbi F (2005) Risk aversion and expected utility theory: A field experiment with large and small stakes. unpublished manuscript
- Buchan N, Croson R (2004) The boundaries of trust: own and others' actions in the US and China. *J Econ Behav Organ* 55:483–504
- Camerer C (1990) Behavioral game theory. In: Hogarth R (ed) *Insights in decision making*. University of Chicago Press, Chicago
- Camerer C (2003) *Behavioural game theory: experiments in strategic interaction*. Princeton University Press, Princeton
- Camerer C (2010) Removing financial incentives demotivates the brain. *Proc Natl Acad Sci* 107:20849–20850
- Camerer C, Hogarth RM (1999) The effects of financial incentives in experiments: a review and capital–labor–production framework. *J Risk Uncertain* 19:7–42
- Camerer C, Nave G, Smith A (2015) Dynamic unstructured bargaining with private information and deadlines: theory and experiment. Caltech HSS working paper
- Cameron LA (1999) Raising the stakes in the ultimatum game: experimental evidence from Indonesia. *Econ Inq* 37:47–59
- Carpenter J, Verhoogen E, Burks S (2005) The effect of stakes in distribution experiments. *Econ Lett* 86:393–398
- Centorrino S, Djemai E, Hopfensitz A, Milinski M, Seabright P (2015) Honest signalling in trust interactions: smiles rated as genuine induce trust and signal higher earning opportunities. *Evol Hum Behav* 36:8–16
- Cherry TL, Frykblom P, Shogren J (2002) Hardnose the dictator. *Am Econ Rev* 92:1218–1221
- Christie R, Geis F (1970) *Studies in Machiavellianism*. Academic Press, New York
- Cooper DJ, Dutcher EG (2011) The dynamics of responder behavior in ultimatum games: a meta-study. *Exp Econ* 14:519–546
- Dalbert C (1999) The world is more just for me than generally: about the personal belief in a just world scale's validity. *Soc Justice Res* 12:79–98

- Dalbert C, Montada L, Schmitt M (1987) Glaube an eine gerechte welt als motiv: Validierungskorrelate zweier Skalen. *Psychol Beitr* 29:596–615
- Dickinson DL (2000) Ultimatum decision-making: a test of reciprocal kindness. *Theor Decis* 48:151–177
- Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner G (2011) Individual risk attitudes: measurement, determinants, and behavioural consequences. *J Eur Econ Assoc* 9:522–550
- Engel C (2011) Dictator games: a meta study. *Exp Econ* 14:583–610
- Falk A, Fehr E (2003) Why labour market experiments? *Labour Econ* 10:399–406
- Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Q J Econ* 114:817–868
- Fehr E, Fischbacher U, Tougareva E (2014) Do high stakes and competition undermine fair behaviour? Evidence from Russia. *J Econ Behav Organ* 108:354–363
- Fehr-Duda H, Bruhin A, Epper TF, Schubert R (2010) Rationality on the rise: why relative risk aversion increases with the stake-size. *J Risk Uncertain* 40:147–180
- Fetchenhauer D, Dunning D (2009) Do people trust too much or too little? *J Econ Psychol* 30:263–276
- Forsythe R, Horowitz JL, Savin NE, Sefton M (1994) Fairness in simple bargaining games. *Games Econ Behav* 6:347–369
- Fouraker L, Siegel S (1963) *Bargaining and group decision making*. McGraw-Hill, New York
- Fu T-T, Kong W-H, Yang CC (2007) Monetary stakes and socioeconomic characteristics in ultimatum games: an experiment with nation-wide representative subjects. Working paper
- Gehrig T, Güth W, Levati V, Levínski R, Ockenfels A, Uske T, Weiland T (2007) Buying a pig in a poke: an experimental study of unconditional veto power. *J Econ Psychol* 28:692–703
- Gillis MT, Hettler PL (2007) Hypothetical and real incentives in the ultimatum game and Andreoni's public goods game: an experimental study. *Eastern Econ J* 33:491–510
- Glaeser E, Laibson D, Scheinkman JA, Soutter CL (2000) Measuring trust. *Q J Econ* 115:811–846
- Goldreich D, Pomorski L (2011) Initiating bargaining. *Rev Econ Stud* 78:1299–1328
- Griffin J, Nickerson D, Wozniak A (2012) Racial differences in inequality aversion: evidence from real world respondents in the ultimatum game. *J Econ Behav Organ* 84:600–617
- Guala F (2005) *The methodology of experimental economics*. Cambridge University Press, New York
- Güth W (2012) Bargaining and negotiations: what should experimentalists explore more thoroughly? In: Croson RT, Bolton GE (eds) *Oxford handbook of economic conflict resolution*, chap 17. Oxford University Press, New York, pp 241–253
- Güth W, Tietz R (1987) Ultimatum bargaining for a shrinking cake: an experimental analysis. In: Tietz R, Albers W, Selten R (eds) *Bounded rational behaviour in experimental games and markets*. Springer, Berlin
- Güth W, Kirchkamp O (2012) Will you accept without knowing what? The Yes–No game in the newspaper and in the lab. *Exp Econ* 15:656–666
- Güth W, Schmittberger R, Schwarze B (1982) An experimental analysis of ultimatum bargaining. *J Econ Behav Organ* 3:367–388
- Güth W, Schmidt C, Sutter M (2007) Bargaining outside the lab—a newspaper experiment of a three-person ultimatum game. *Econ J* 117:449–469
- Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R (2001) In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am Econ Rev* 91:73–78
- Herrmann B, Thöni C, Gächter S (2008) Antisocial punishment across societies. *Science* 319:1362–1367
- Hertwig R, Ortmann A (2001) Experimental practices in economics: a methodological challenge for psychologists? *Behav Brain Sci* 24:383–451
- Ho T-H, Weigelt K (2005) Trust building among strangers. *Manag Sci* 51:519–530
- Hoffman E, McKabe K, Schachat K, Smith V (1994) Preferences, property rights, and anonymity in bargaining. *Games Econ Behav* 7:346–380
- Hoffman E, McKabe K, Smith V (1996) On expectations and the monetary stakes in ultimatum games. *Int J Game Theory* 25:289–301
- Holm HJ, Nystedt P (2008) Trust in surveys and games—a methodological contribution on the influence of money and location. *J Econ Psychol* 29:522–542
- Holt CA, Laury SK (2002) Risk aversion and incentive effects. *Am Econ Rev* 92:1644–1655
- Johansson-Stenman O, Mahmud M, Martinsson P (2005) Does stake size matter in trust games. *Econ Lett* 88:365–369
- Johansson-Stenman O, Mahmud M, Martinsson P (2013) Trust, trust games and stated trust: evidence from rural Bangladesh. *J Econ Behav Organ* 95:286–298
- Johnson ND, Mislin AA (2011) Trust games: a meta-analysis. *J Econ Psychol* 32:865–889

- Kachelmeier SJ, Shehata M (1992) Examining risk preferences under high monetary incentives: experimental evidence from People's Republic of China. *Am Econ Rev* 82:1120–1141
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291
- Kang MJ, Rangel A, Camus M, Camerer C (2011) Hypothetical and real choice differentially activate common valuation areas. *J Neurosci* 31:461–468
- Karagözoğlu E (2012) Bargaining games with joint production. In: Croson RT, Bolton GE (eds) *Oxford handbook of economic conflict resolution*, chap 24. Oxford University Press, New York, pp 359–372
- Karagözoğlu E, Riedl A (2015) Performance information, production uncertainty, and subjective entitlements in bargaining. *Manag Sci* 61:2611–2626
- Kocher MG, Martinsson P, Visser M (2008) Does stake-size matter for cooperation and punishment? *Econ Lett* 99:508–511
- Konow J (2005) Blind spots: the effects of information and stakes on fairness bias and dispersion. *Soc Justice Res* 18:349–390
- Kühberger A (2001) Why use real and hypothetical payoffs? *Behav Brain Sci* 24:419–420
- Leibbrandt A, Maitra P, Neelim A (2015) On the redistribution of wealth in a developing country: experimental evidence on stake and framing effects. *J Econ Behav Organ* 118:360–371
- Levitt SD, List JA (2007) What do laboratory experiments measuring social preferences reveal about the real world? *J Econ Perspect* 21:153–174
- List JA (2007) On the interpretation of giving in dictator games. *J Polit Econ* 115:482–493
- List JA, Cherry TL (2000) Learning to accept in ultimatum Games: evidence from an experimental design that generates low offers. *Exp Econ* 3:11–29
- List JA, Cherry TL (2008) Examining the role of fairness in high stakes allocation decisions. *J Econ Behav Organ* 65:1–8
- Markowitz H (1952) The utility of wealth. *J Polit Econ* 60:151–158
- McCrae RR, Costa PT, Martin TA (2005) The NEO-PI-3: a more readable revised NEO personality inventory. *J Personal Assess* 84:261–270
- McGuire JT, Botvinick MM (2010) Prefrontal cortex, cognitive control, and the registration of decision costs. *Proc Natl Acad Sci* 107:7922–7926
- Menezes CF, Hanson DL (1970) On the theory of risk aversion. *Int Econ Rev* 11:481–487
- Munier B, Zaharia C (2003) High stakes and acceptance behavior in ultimatum bargaining. *Theor Decis* 53:187–207
- Murayama K, Matsumoto M, Izuma K, Matsumoto K (2010) Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proc Natl Acad Sci* 107:20911–20916
- Myerson R (1979) Incentive compatibility and the bargaining problem. *Econometrica* 47:61–73
- Myerson R (1984) Two-person bargaining problems with incomplete information. *Econometrica* 52:461–488
- Naef M, Schupp J (2009) Measuring trust: experiments and surveys in contrast and combination. IZA discussion paper series #4087
- Nagel R, Mitzkewitz M (1993) Experimental results on ultimatum games with incomplete information. *Int J Game Theory* 22:171–198
- Neelin J, Sonnenschein H, Spiegel M (1988) A further test of noncooperative bargaining theory. *Am Econ Rev* 78:824–836
- Oberholzer-Gee F, Waldfogel J, White MW (2010) Friend or foe? Cooperation and learning in high-stakes games. *Rev Econ Stat* 92:179–187
- Oosterbeek H, Sloof R, van de Kuilen G (2004) Cultural differences in ultimatum game experiments: evidence from a meta-analysis. *Exp Econ* 7:171–188
- Parco JE, Rapoport A, Stein WE (2002) Effects of financial incentives on the breakdown of mutual trust. *Psychol Sci* 13:292–297
- Post T, van den Assem M, Baltussen G, Thaler R (2008) Deal or no deal? Decision making under risk in a large-payoff game show. *Am Econ Rev* 98:38–71
- Raihani NJ, Mace R, Lamba S (2013) The effect of \$1, \$5, and \$10 stakes in an online dictator game. *Plos ONE* 8:e73131
- Rabin M (1993) Incorporating fairness into game theory and economics. *Am Econ Rev* 83:1281–1302
- Read D (2005) Monetary incentives, what are they good for? *J Econ Methodol* 12:265–276
- Roth AE, Prasnikar V, Okuno-Fujivara M, Zamir S (1991) Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. *Am Econ Rev* 81:1068–1095

- Rubin Z, Peplau LA (1975) Who believes in a just world? *J Soc Issues* 31:65–89
- Rubinstein A (1982) Perfect equilibrium in a bargaining model. *Econometrica* 50:97–110
- Schmitt M (1996) Individual differences in sensitivity to befallen injustice. *Person Individ Differ* 21:3–20
- Sefton M (1992) Incentives in simple bargaining games. *J Econ Psychol* 13:263–276
- Slonim R, Roth AE (1998) Learning in high stakes ultimatum games: an experiment in the Slovak Republic. *Econometrica* 66:569–596
- Smith V, Walker JM (1993) Monetary rewards and decision costs in experimental economics. *Econ Inq* 31:245–261
- Sonnegård J (1996) Determination of first movers in sequential bargaining games: an experimental study. *J Econ Psychol* 17:359–386
- Ståhl S (1972) Bargaining theory. Economic Research Institute, Stockholm School of Economics, Stockholm
- Straub PG, Murnighan JK (1995) An experimental investigation of ultimatum games: information, fairness, expectations, and lowest acceptable offers. *J Econ Behav Organ* 27:345–364
- Sutter M, Kocher MG (2007) Trust and trustworthiness across different age groups. *Games Econ Behav* 59:364–382
- Telser KG (1993) The ultimatum game: a comment. Mimeo, University of Chicago
- Thaler RH (1988) Anomalies: the ultimatum game. *J Econ Perspect* 2:195–209
- Tompkinson P, Bethwaite J (1995) The ultimatum game: raising the stakes. *J Econ Behav Organ* 27:439–451
- Van den Assem MJ, van Dolder D, Thaler R (2012) Split or steal? Cooperative behavior when the stakes are large. *Manag Sci* 58:2–20
- Vieider FM (2012) Moderate stake variations for risk and uncertainty, gains and losses: methodological implications for comparative studies. *Econ Lett* 117:718–721
- Zhou Y, Wang Y, Rao L-L, Yang L-Q, Li S (2014) Money talks: neural substrate of modulation of fairness by monetary incentives. *Front Behav Neurosci* 8:1–11
- Zwicker R, Chen X-P (1999) What price fairness? A bargaining study. *Manag Sci* 45:804–823