The Cost of Punishment: Fairness in High-Stakes Ultimatum Games

Kent Van Donge
Faculty Advisor: Daniel Acland

May 1, 2015

Abstract: This paper presents an expanded experimental design of the standard ultimatum game in order to understand how perceptions of fairness change with changes in monetary stakes. Previous experiments have been unable to independently vary either the utility of punishing an unfair offer or the monetary cost of rejecting said offer. By introducing an exogenous cost to rejecting an offer, it is possible to hold one of these terms constant while changing the other. In this way it is possible to gather data that would show how the utility of punishment – and thus, perceptions of fairness – changes when different amounts of money are involved.

I would like to thank my advisor, Daniel Acland, and good friend, Jacob F. Grant, for their helpful comments and critiques throughout this process. Without them this would not have been possible.

Introduction

Ultimatum games are designed to measure how non-monetary considerations impact decision making. In particular, they demonstrate the willingness of an individual to forgo real monetary reward in order to punish an action he or she perceives as unfair. In these games there are two players, a proposer and a responder. The proposer is given an amount of money which he or she must split between the responder and him or herself. After the proposer suggests a split of the total money, the responder can either accept the offer or reject the offer. In the former case both players receive the agreed upon amounts, while in the latter both players receive nothing. Classical economic models that assume rational self-interest suggest that the responder will accept any amount of money offered, as the choice is between some money or no money, and thus the proposer will offer as little money as possible. However, experimental results consistently find that responders will actually reject offers they consider too unfair and proposers will almost always offer relatively fair amounts. These counterintuitive results have inspired a multitude of experiments on this subject and numerous theories on why and how these results occur.

While the theoretical underpinning of ultimatum bargaining games has been around even longer, the first experimental version of the ultimatum game was carried out in 1982 (Guth, Schmittberger, and Schwarze, 1982). Since then, many economic experiments based around ultimatum bargaining have been performed, resulting in a number of different conclusions. An important and oft explored question is how higher monetary stakes affect rejection rates among responders. Experiments involving relatively high monetary stakes have seen substantial variation in their results. Some researchers have suggested that their data demonstrates no significant change in rejection rates (Munier and Zaharia, 2003; Cameron, 1995). However, this contradicts the intuition that as the cost of punishing an offer increases, the willingness to do so decreases. Indeed, many other researchers have found that as stakes increase, rejections do in fact go down (Andersen, Ertac, Gneezy, and List, 2011; Slonim and

Roth, 1998; Hoffman, McCabe and Smith, 1996). There are two reasons that researchers found no significant change in rejection rates with higher monetary stakes: the stakes were not raised high enough to see a substantial effect and the difficulty with getting the proposers to offer low amounts to the responders. Researchers who found no significant change in rejection rates with higher monetary stakes may not have raised the stakes high enough to see a substantial effect or have had difficulty in getting proposers to offer low amounts to the responders. Experiments that increase stakes more dramatically and find ways to generate low offers typically yield the intuitively predictable result: as the total money involved increases the willingness to reject an unfair offer decreases (Andersen, Ertac, Gneezy, and List, 2011).

Some have interpreted this as evidence that as the stakes increase, there is a decrease in demand for punishment – i.e. there is a decrease in the willingness to pay (WTP) for punishment, the monetary value of the utility gained by punishing, which is assumed to be determined by perceptions of fairness (Andersen, Ertac, Gneezy, and List, 2011). However, this cognitive leap is unjustified. Whether or not an offer is rejected is determined by both the behavioral feeling of rejecting an offer (i.e. the utility of punishment) and the monetary cost of rejecting the offer. If the utility gained from the feeling of rejecting an offer is greater than the monetary cost, the responder will reject. Previous experiments have been unable to separate these two components, and thus cannot suggest that a change in rejection rates results from one or the other. Both the monetary cost and the utility of rejecting the offer are determined by the same inputs – total monetary stakes and the percentage offered – and cannot be varied independently. This means that prior experiments cannot reasonably claim to know how the demand for punishing an unfair offer changes with respect to total stakes, except that it must either go down as the stakes go up or increase less than the increase in the monetary cost of rejecting the offer. It is not possible to say with certainty whether the demand for punishing unfair offers decreases, increases, or remains unchanged, because the cost is increasing with increasing stakes, which would, on its own, decrease rejection rates.

The problem of being unable to determine how the demand for punishment is changing arises because it is not possible to vary the monetary cost of rejecting an offer without also varying either the responder's share, the size of the total stakes, or both. To identify the effect of the size of total stakes on the utility of punishing an unfair offer you have to be able to vary the total monetary cost of rejecting without simultaneously varying the utility of punishment. I am proposing an experimental design that will make this possible by adding an exogenous cost to rejecting an offer that can be varied by the experimenter. Instead of neither party receiving any money if the offer is rejected there is an extra fee charged to the responder for rejecting the offer, thus making it possible to change the cost of rejecting without changing the total stakes or the percentage of the money offered to the responder. This extra cost allows the utility of punishment to be held constant while cost is changed, and vice versa. If we think of the percentage of the total stakes offered to the responder as a determinant of the magnitude of punishment utility gained by rejection, then, in terms of consumer choice theory, my design allows the quantity and price of punishment to be varied independently. Thus it becomes possible to experimentally determine the effect of varying the utility of punishing on rejection rates independent of the cost and then to identify the effect of the size of the stakes on the utility of punishment, i.e. the perception of unfairness, independent of its effect on the cost.

Model

Previous experimental designs have been unable to determine the exact nature of the relationship between the total stakes and the utility of punishment because varying the total stakes changes both the utility derived from the punishing an unfair offer and the monetary cost of rejecting said offer. It has been impossible to vary one without varying the other. The experiment and model I am proposing solves this problem by introducing an exogenous cost, imposed by the experimenter, which will allow the experimenter to isolate the effect of changing the total stakes on the utility of punishing, while holding the cost of

rejection constant.

Under the assumption that the decision to reject is determined purely by the utility from punishing and the cost of doing so, the utility function for a respondent considering an offer in a standard ultimatum game, assuming quasi-linear utility of money, looks like this:

$$\begin{aligned} U(R;\rho,T) &= R \cdot g(\rho,T) + (1-R)\cdot \rho T \\ &= \begin{cases} g(\rho,T) & \text{if } R=1 \\ \rho T & \text{if } R=0 \end{cases} \end{aligned}$$

Where

U =Utility of Responder
R =Bernoulli Variable for Accept or Reject
($R=1$ when the offer is rejected and $R=0$ when the offer is accepted)
g =Utility of Punishing
ρ =Proportion of the Total Money Offered to the Responder
T =Total Monetary Stakes Begin Split

The indifference point between accepted and rejected offers, ρ^*, and how it varies with changes in the total stakes, can be observed with the standard ultimatum game experiment. It is simply the point that, for a given level of money being split, the responder is indifferent between rejecting or accepting the offer. The decision rule is:

$$R = 1 \; if \; g(\rho,T) > \rho T$$

Thus ρ^* satisfies

$$\begin{aligned} g(\rho,T) &= \rho^* T \\ \Rightarrow \rho^* &= \frac{g(\rho,T)}{T} \end{aligned}$$

Taking the derivative of ρ^* with respect to T we get:

$$\frac{\partial \rho^*}{\partial T} = \frac{\frac{\partial g}{\partial T}}{T} - \frac{g}{T^2}$$

This is the object identified in the existing literature on changing stakes. However, this is

not actually very useful in understanding how people's perception of fairness changes when the total amount of money in play changes. That change is $\frac{\partial g}{\partial T}$, and cannot be determined with the gathered data. Whenever ρ is varied to identify ρ^*, and T is subsequently varied to identify $\frac{\partial \rho^*}{\partial T}$, the value of the utility of punishment, $g(\rho, T)$, also changes. Since we are unable to hold $g(\rho, T)$ constant while varying either ρ or T we cannot observe the effects of changing stakes on the utility of punishment and the cost of rejecting independently. For a given ρ and T the only thing that can be observed is whether the offer is accepted or rejected, which could be due to either the monetary cost or the utility of punishment changing.

This problem can be solved by imposing an additional external cost to rejecting the offer. The exogenous cost can be determined at the discretion of the experimenter. This means that the total monetary cost of rejecting an offer can be varied without changing either ρ or T, or conversely, the monetary cost can be held constant while varying ρ and T. I add this new external cost parameter to the classical model below:

$$\begin{aligned} U(R; \rho, T, K) &= R \cdot g(\rho, T) + (1 - R) \cdot \rho T - R \cdot K \\ &= \begin{cases} g(\rho, T) - K & \text{if } R = 1 \\ \rho T & \text{if } R = 0 \end{cases} \end{aligned}$$

Where

$$K = \text{Exogenous Fee for Rejecting}$$

We can define the new total cost of rejecting as $h = \rho T + K$ and by varying K to find the indifference point $h^*(\rho, T, K)$ we can observe the willingness to pay for punishment, which is the monetary value of the utility of rejection, $g(\rho, T)$. Thus, the change in the indifference point $h^*(\rho, T, K)$ over varying levels of T is exactly equal to the change in $g(\rho, T)$. The decision rule becomes:

$$\begin{aligned} R = 1 \; if \; & g(\rho, T) - K > \rho T \\ & \Rightarrow g(\rho, T) > \rho T + K \\ & \Rightarrow g(\rho, T) > h \end{aligned}$$

The indifference point satisfies

$$h^*(\rho, T) = g(\rho, T)$$

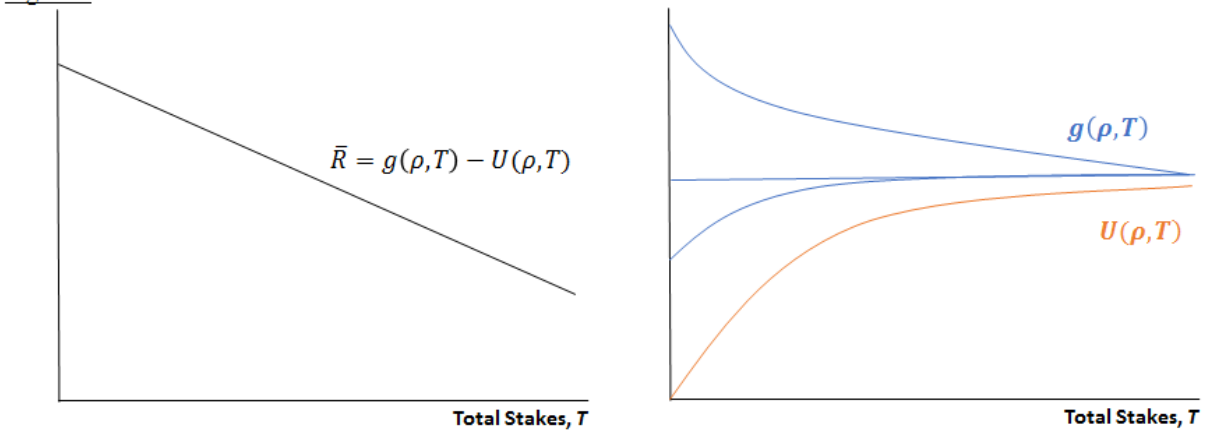
Therefore

$$\frac{\partial h^*}{\partial T} = \frac{\partial g}{\partial T}$$

By using varying levels of K to find the indifference point $h^*(\rho, T, K)$ the total monetary cost of rejection can be varied while holding both ρ and T constant. Thus any change in $h^*(, T, K)$ is due only to changes in the utility of punishment.

Prior experiments show that responder demand for rejecting offers decreases as the total stakes increase (Andersen, Ertac, Gneezy, and List, 2011). Thus, while it may seem logical to conclude that responder demand for punishing proposers is also decreasing, this conclusion is not justified. The rejection rate, $\bar{R}$, is a function of the difference between the utility of punishment and the utility of the monetary cost, $g(\rho, T) - U(\rho, T)$. Under the assumption that $\frac{\partial U(\rho, T)}{\partial T} > 0$, $\bar{R}$ will be decreasing, as depicted in the first panel of figure 1, whenever $\frac{\partial g(\rho, T)}{\partial T} < \frac{\partial U(\rho, T)}{\partial T}$. Three shapes for $g(\rho, T)$ that are consistent with previously observed results on the effect of stakes on rejection rates are depicted in the second panel of figure 1. Previously it was impossible to determine not only the value of $\frac{\partial g}{\partial T}$, but even whether it was positive or negative. With the model and experiment I propose, this will be possible.

Figure 1



Experimental Design

Hypothetically, an experiment can consist of a group of volunteers being randomly split into responders and proposers. In reality though, the proposers are actually confederates posing as proposers. We would present the game and its rules to the participants and have the responders respond to different, predetermined offers in an ultimatum game, then either give them the agreed upon amount or charge them the predetermined cost if they reject. The subjects would be paid a base amount as well for simply participating in the experiment, and the externally imposed cost, K, would be removed from their payment instead of being charged directly to any responder who rejects the offer. The points at which these responders are indifferent between accepting and rejecting the offer are most significant. This data will be gathered by varying amounts of K while holding T and ρ constant until this indifference point is reached. The process can then be repeated for various levels of T. For this reason, a calibration run would be necessary prior to the final experiment in order to determine the approximate range in which indifference point will be located. Depending on how precisely an experimenter wanted to measure the indifference point, more calibration runs might be necessary. These indifference points are equal to the utility of punishment at a specific value of ρ and T and thus the data will provide insight into how the utility of punishing, and therefore perceptions of fairness, changes with monetary stakes.

The goal of this experiment would be to determine how the utility of punishing an unfair offer fluctuates with changes in total stakes. This makes the behavior of proposers insignificant in this particular experiment. Also, it is necessary to control for and keep constant the percentage offered to responders by proposers in order for the data to be viable. For these reasons it is not only acceptable, but also necessary for the experimenters to deceive the responders by posing as the proposers. In this way the amount offered to the proposer can be exactly controlled for and manipulated. The only alternative would be to play so many variations of the game that there would be enough of the exact percentage

values offered naturally to yield statistically significant data. This is not realistically feasible. Another problem that arises is in general subjects will not offer very unfair offers at high stakes. An ongoing debate within experimental economics centers on both the ethics and the practicality of using deception in experiments. There have been studies questioning this proscription and purporting that deception is not adverse in its effects on experimental results (Croson, 2005; Weimann, 1992; Bonetti, 1998). In fact, there has even been a study in neuroeconomics which involved scanning a subject's brain as he or she was deceived while playing an ultimatum game and found no evidence that deception impacted the results (Sanfey, Rilling, Aronson, Nystrom, and Cohen, 2003). As a caveat though, this deception can have an impact in games being played multiple times, but since this experiment is being played only once this is not an issue.

It is possible that when imposing the external fee for rejecting an offer there may be an issue with loss aversion. The monetary cost of receiving no money from rejecting versus paying extra money to reject may not be viewed in the same way, and thus may make it untenable to correctly hold the monetary cost to the responder constant. The solution is to make the extra fee the subjects have to pay come out of whatever they are being paid for taking the experiment. Instead of having the responders pay a fee in order to reject the offer, they simply have to give up a certain amount of the money they are being paid to participate in the experiment. As a result both the money they give up and the fee they pay for rejecting an offer are both in the domain of reducing gains, instead of one reducing gains and the other incurring losses.

Finally, there are the additional problems that plague standard ultimatum games: correct wording of the question so as not to confuse the subjects, making the ultimatum game believable so subjects will act in accordance with their true preferences, and having proposers offer splits that are not close to an equal split. The problem of requiring an unequal split has already been addressed and solved by having experimenters pose as proposers so

they can be instructed exactly what to offer. The problem of volunteers trying to game the system, rather than act as they would in a real-,life scenario is solved in the same way as many other experimenters, by playing the game with real stakes and actually paying the people whatever they end up agreeing on at the end of the experiment. Ultimately, making sure that the subjects understand how the game works and ensuring they are not confused is an important issue, particularly for ultimatum games. Luckily many experimenters have also faced this problem and have already come up with solutions. For this reason, it would likely be best to borrow the wording from an experiment that has already worked and modify it to fit this particular experiment.

Interpretation of Possible Data & Concluding Remarks

The design of previous ultimatum game experiments did not include the ability to gather data about how the utility of punishing changes with changes in total stakes. Recent experiments have concluded that the rejection rates in ultimatum games decrease as the total monetary stakes involved increase. While this may seem to imply that the utility of punishing becomes less important as the monetary cost of rejection increases, meaning there would be a downward sloping demand curve for punishing unfair offers, in reality this is not the case. The only thing this tells us is that the derivative of the utility of punishing with respect to total stakes is less than the derivative of the total cost of rejecting the offer with respect to total stakes. It cannot tell us the sign of the derivative of the utility of punishing, which is what is significant in trying to gain an insight into how people perceive and react to fairness. Also, while lower rejection rates with higher stakes have been observed in many recent papers, there are still a multitude of experiments that purport no significant change in rejection rates with changes in stakes. A common explanation for these results is that either the stakes were not raised high enough for the change to be noticeable or because of the difficulty in eliciting enough low proportion offers from proposers in natural ultimatum games. While these explanations are intuitively reasonable there is still a possibility that these results might be better explained by how the utility of punishing changes with stakes.

How the utility of punishment, and thus people's perception of fairness, changes when more or less money is involved is still unknown. It is possible that the demand for punishment as a function of the stakes is either downward or upward sloping. A negative value of $\frac{\partial g}{\partial T}$ indicates a downward sloping demand for punishment. As the monetary stakes increase the utility derived from punishing an offer of relatively the same unfairness decreases. But a positive value of $\frac{\partial g}{\partial T}$ could also explain the same results. As total stakes increase the utility for punishing an offer, with the share offered being held constant, increases. This implies an upward sloping demand for punishment as a function of stakes, indicating that people perceive a given offer as less fair the more money is at stake. With the data currently available, both explanations remain equally plausible.

Current ultimatum game experiments cannot determine how the utility of punishing an unfair offer will change when the stakes are changed. We can only observe whether someone rejects or accepts an offer, but previous experiments have been unable to hold only the utility of punishment or the monetary cost of rejecting an offer constant while varying the other. This makes it impossible to determine just the utility of punishment and how it changes as stakes change. By introducing an exogenous cost variable into the equation of whether or not a responder will reject or accept an offer, it is possible to hold the total stakes and percentage offered, and thus the utility of punishment, constant while varying the monetary cost of rejecting said offer. It is possible to determine the responder's utility of punishment in monetary terms, as a result of the introduction of the exogenous cost. Thus, it is possible to determine how this utility of punishment, and therefore the perception of fairness, will change as other variables change.

References

Andersen, Steffen, Seda Ertac, Uri Gneezy, Moshe Hoffman, and John A. List. 'Stakes Matter In Ultimatum Games'. *American Economic Review* 101.7 (2011): 3427-3439.

Bonetti, Shane. 'Experimental Economics And Deception'. *Journal of Economic Psychology* 19.3 (1998): 377-395.

Cameron, Lisa A. 'RAISING THE STAKES IN THE ULTIMATUM GAME: EXPERIMENTAL EVIDENCE FROM INDONESIA'. *Economic Inquiry* 37.1 (1999): 47-59.

Croson, Rachel. 'The Method Of Experimental Economics'. *International Negotiation* 10.1 (2005): 131-148.

Gth, Werner, Rolf Schmittberger, and Bernd Schwarze. 'An Experimental Analysis Of Ultimatum Bargaining'. *Journal of Economic Behavior & Organization* 3.4 (1982): 367-388.

Hoffman, Elizabeth, Kevin A. McCabe, and Vernon L. Smith. 'On Expectations And The Monetary Stakes In Ultimatum Games'. *International Journal of Game Theory* 25.3 (1996): 289-301.

Munier, Bertrand, and Costin Zaharia. 2002. "High Stakes and Acceptance Behavior in Ultimatum Bargaining: A Contribution from an International Experiment." Theory and Decision, 53(3): 187- 207

Sanfey, A. G. 'The Neural Basis Of Economic Decision-Making In The Ultimatum Game'. *Science* 300.5626 (2003): 1755-1758.

Slonim, Robert, and Alvin E. Roth. 'Learning In High Stakes Ultimatum Games: An Experiment In The Slovak Republic'. *Econometrica* 66.3 (1998): 569.

Weimann, Joachim. 'Individual Behaviour In A Free Riding Experiment'. *Journal of Public Economics* 54.2 (1994): 185-200.