Rotemberg, Julio J.

Working Paper

Minimally acceptable altruism and the ultimatum game

Working Papers, No. 06-12

Provided in Cooperation with:
Federal Reserve Bank of Boston

Suggested Citation: Rotemberg, Julio J. (2006) : Minimally acceptable altruism and the ultimatum game, Working Papers, No. 06-12, Federal Reserve Bank of Boston, Boston, MA

This Version is available at:
https://hdl.handle.net/10419/55606

Working Papers
Federal Reserve Bank of Boston

No. 06-12

Minimally Acceptable Altruism and the Ultimatum Game

Julio J. Rotemberg

Abstract:
I suppose that people react with anger when others show themselves not to be minimally altruistic. With heterogeneous agents, this can account for the experimental results of ultimatum and dictator games. Moreover, it can account for the surprisingly large fraction of individuals who offer an even split, with parameter values that are more plausible than those required to explain outcomes in these experiments with the models of Levine (1998), Fehr and Schmidt (1999), Dickinson (2000), and Bolton and Ockenfels (2000).

JEL Classifications: D64, D63, C72, A13

Julio J. Rotemberg is William Ziegler Professor of Business Administration at the Harvard Business School and a visiting scholar at the Federal Reserve Bank of Boston. His email address is jrotemberg@hbs.edu.

I wish to thank Max Bazerman, Rafael Di Tella, David Laibson, and David Scharfstein for helpful conversations, and the Harvard Business School Division of Research for research support.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at http://www.bos.frb.org/economic/wp/index.htm.

This version: May 23, 2006

This paper presents a model of individual preferences where individuals are mildly altruistic towards others while also expecting others to be mildly altruistic. If an individual encounters evidence that another is less altruistic than he finds acceptable, he becomes angry and derives pleasure from harming the excessively selfish individual. These preferences are shown to be capable of explaining the experimental outcomes of the ultimatum game of Güth *et al.* (1982) as well as those of an important variant, namely the dictator game of Forsythe *et al.* (1994). The main advantage of the model proposed here is that, unlike other models of social preferences that have been proposed to explain these experiments, it does not lead individuals to take unrealistic actions outside these experimental settings.

Because the experimental outcomes of ultimatum and dictator games are in such sharp conflict with the predictions of standard economic models, many experimental variations have been considered and this experimental literature is vast. Still, it is worth recalling the settings and some of the findings. Both games involve two players. The first player, who is called the proposer, offers to split a pie with the second, who is called the responder. In the ultimatum game, the responder can either accept or reject the proposer's offer. If the responder rejects it, neither player gets anything. Otherwise, the pie is split in the way suggested by the proposer. In the dictator game, the responder must passively accept the proposer's offer.

The modal offer in the ultimatum game is to split the pie 50-50. The actual fraction of even splits varies somewhat from experiment to experiment, and sometimes varies across rounds of play within a given experiment. In Forsythe *et al.* (1994), about half the proposers offer an even split while Levine (1998) reports that about 28 percent of proposers offered an even split in the late rounds of the Roth *et al.* (1991) experiments. Splits that are less favorable to responders often get rejected. In several experiments (see Figure 6 of Roth *et al.* (1991) and Harrison and McCabe (1996)), such rejections are so common that average earnings of proposers actually decline as they make offers that are less favorable to responders. With this behavior of responders, proposers should offer even splits if they wish to maximize their own expected payoffs. However, even in experiments where proposers earn

less by making less generous offers and even after subjects have learned the game by playing it several times, some proposers make less generous offers. [1] In the dictator game, expected monetary payoffs obviously rise when less generous offers are made. Not surprisingly, this implies that offers of even splits are observed less frequently. Still, Forsythe *et al.* (1994) report that about 20 percent of their proposers in the dictator game offered even splits.

The model presented below is closely related to Levine (1998) who also supposes that agents' altruism for others depends on their assessment of how altruistic others are in return. Unlike Levine (1998), an agent's altruism is not assumed to depend linearly on that agent's perception of the altruism of the person he is interacting with. This allows the model to avoid Levine's (1998) conclusion that most people derive pleasure from seeing others suffer. His baseline parameters imply that 72 percent of the population would be willing to give up more than 25 cents to ensure that a stranger loses a dollar while 20 percent are ready to give up over 95 cents to bring about this outcome. In his model, this nastiness (or spitefulness) is important because it explains both why many responders reject uneven offers and why some proposers reduce their expected earnings by making offers that are less generous than even splits.

One difficulty with the preferences implied by Levine's (1998) analysis is that, given the limitations of actual law enforcement institutions, they ought to lead to massive amounts of vandalism. The experimental evidence in other games also casts doubt on the ubiquitousness of spite. Using variants of dictator games, Charness and Rabin (2002) show that most people would actually sacrifice some of their own resources to induce small gains by others whereas essentially no one is willing to sacrifice significantly to hurt others. They show, in particular, that 73 percent of their subjects are willing to give up 100 of their units to cause another agent to gain 400. Not only would all the agents in Levine (1998) turn this down (because even his altruists are not sufficiently altruistic for this) but 20 percent of his agents prefer

[1] It should be noted, however, that in most of the ultimatum experiments in primitive societies reported by Henrich *et al.* (2004 p. 25), the offer that maximizes the income of the proposer is usually below both the mean offer and the 50-50 split. Thus, in these societies, the even splits that are commonly observed constitute generous offers.

an outcome where they and another player both receive nothing to an outcome where the other player receives 80 while they get 20. In the Charness and Rabin (2002) experiments, almost all the players choose the 80/20 outcome over the one where neither gets anything.[2]

In an Appendix, Charness and Rabin (2002) present a model where agents who have acted in ways that are not consistent with maximizing social preferences accumulate "demerits" and where these demerits make agents less liked by others. While it features agents whose preferences are all identical, this model is similar in spirit to Levine (1998) and to the present paper. However, Charness and Rabin (2002, p. 851) make it clear that, in part because of the assumption of homogeneous preferences, which rules out different proposers making different offers or different responders responding differently, they do not intend this model to be "useful in its current form for calibrating experimental data."[3]

Many of the existing models of fairness that are spelled out in sufficient detail so that it is possible to know the parameter values they require to explain laboratory phenomena do not explicitly suppose that individual utilities depend on the parameters of other agents' utility functions. They suppose instead that people maximize a function that depends on the payoffs of several agents. There are two related ways in which this has been done. In the pioneering paper of Rabin (1993), an agent regards another as fair if he expects his actions to be "kind," where kindness of the second player towards the first is defined as the difference between the payoffs that the first expects to receive from the second and the payoffs that it would have been "equitable" for the second to give to the first. Rabin (1993) then defines equitable payoffs as the average of the highest and the lowest payoffs the second agent can

[2]As I discuss below, the generous behavior observed in any dictator game, including those of Charness and Rabin (2002) could be due to fear of being found out and punished. However, relatively large punishments are required for the generosity observed by Charness and Rabin (2002) to be consistent with the spitefulness assumed by Levine (1998).

[3]Charness and Rabin (2002 p. 857) do prove a theorem for the case where all individuals have a higher level of altruism than the level of altruism they demand of others. They show that, in this case, equilibria where people simply maximize their altruistic utility are also equilibria when people punish those that have accrued demerits. They go on to say that "there may additionally be negative" equilibria where people punish each other because they expect more generosity than is forthcoming in equilibrium. However, even splits in ultimatum games are more "positive" outcomes than those that simply maximize altruistic utility, at least in the realistic case where people care more for a dollar in their own pocket than for a dollar in someone else's pocket.

give to the first under the assumption that the player acts efficiently. Lastly, he supposes that agents maximize the sum of their own material payoffs and the product of their own kindness times the kindness they expect to receive from the other agent.

This model has been extended by both Dickinson (2000) and Falk and Fischbacher (2006) to account for the results of ultimatum games. One important modification in Dickinson (2000) and Falk and Fischbacher (2006) is that they allow an agent's utility function to put lower weight on the agent's own material payoffs than on "reciprocity" *i.e.* on the product of the agent's own kindness and the kindness he receives from the other player. Offers of even splits can be obtained as equilibria in both papers, but only when agents care exclusively about reciprocity and ignore their own material payoffs. For obvious reasons, this limit is not particularly attractive as an empirical description of preferences.

The alternative approach of Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) supposes that agents care about how their own payoffs compare to those of others. These widely cited papers rationalize high offers in the ultimatum game by the unwillingness of responders to accept offers that give them payoffs that are low relative to those received by the proposer. Taken literally, these models imply that rejections do not hinge on the alternate courses of action that were available to the proposer and depend only on how the allocation of resources embodied in the proposer's offer compares to the allocation that results from rejection. However, the results of Kagel and Wolfe (2001) cast doubt on the importance of inequality in leading to rejections. They consider a variant of the ultimatum game where rejection by the responder leads to a payment to a third party - so that inequality can be increased by rejection. In Kagel and Wolfe's (2001) experiments, the rate of rejection does not depend significantly on the amount that this third party receives when the proposer's offer is rejected.

These models are also inconsistent with the evidence showing that the alternatives available to the proposer are an important determinant of responder behavior. Blount (1995), in particular, shows that the actions of responders depend on whether proposers themselves make the offer or whether the offer is randomly chosen by a computer. Similarly, Falk and

Fischbacher (2006) and Charness and Rabin (2002) show that responders' reactions depend not only on the final offer but also on the initial choices available to the proposer. Even leaving this aside, the Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) models seem incapable of explaining the outcomes in ultimatum games except under implausible assumptions concerning parameters. As shown in section 1, these models can predict even splits only when responders are unrealistically mean or when proposers are unrealistically kind hearted (or both).

The model of section 2 supposes instead that, unless they feel provoked by evidence that someone is blameworthy, most people are modestly altruistic towards those around them. In principle, they are thus willing to incur small costs if this leads others to obtain large gains. Individuals are also assumed to care intensely about whether others are altruistic. This is broadly consistent with the evidence that responders care more about the actions of proposers than they do about the allocations that result from acceptance or rejection, because this evidence suggests that the proposer's intentions matter to responders. I suppose in particular that there is a minimal level of altruism that each person expects from those that he interacts with and that demonstrations that one is more selfish than this are met with strong disapproval, and even anger. By contrast, demonstrating a level of altruism above the minimal level has a much smaller effect.[4] The idea that people expect a minimal level of altruism from others fits with the notion that people are expected to be considerate and, more generally, to have manners.[5]

The notion that people change their preferences drastically when they feel that someone's actions demonstrate insufficient altruism is consistent with Pillutla and Murnighan's (1996) evidence that rejections in ultimatum games are associated with anger. One also observes rapid angry responses in the field when people feel mistreated by strangers. These emotions

[4]For examples of nonlinear preferences defined over payoffs (as opposed to over beliefs about preferences) of the two players see, for example, Loewenstein, Thompson and Bazerman (1989).

[5]The expectation that people are at least somewhat altruistic also justifies trust in others. This can lead individuals to rely on the judgment of others about what is good for them. This may, in particular, help to rationalize the finding by Choi *et al.* (2004) that individuals overwhelmingly choose the "default" option when firms offer them a choice of retirement packages.

seem particularly salient in "road rage," the sometimes violent reaction of drivers who feel that other drivers have acted badly.[6]

Anger at insufficient altruism can explain the tendency of responders to reject uneven offers. This leaves the question of how one can rationalize the observation that some proposers actually reduce their expected earnings by making uneven offers. These lower expected earnings also raise questions about the even-handedness of responders, since it would appear that these are punishing proposers whose actions are irrational. It turns out, however, that both the making of low offers and the recurrent rejections of such offers can be rationalized if experimental subjects sometimes act as if they were risk-loving. For a proposer, offers that are less favorable to the responder than an even split constitute a gamble in the sense that such offers are rejected with positive probability. Risk loving proposers with relatively low altruism levels would thus seek such gambles even if the expected returns of these gambles were slightly below those obtained from offering an even split.

For this explanation to be plausible, one must believe that experimental subjects can act in risk-loving ways. As I discuss below, risk-loving behavior has been observed in several different experimental settings. At the same time, it must be recognized that the evidence is mixed in the sense that other experiments have found subjects to be risk-averse.[7]

The paper proceeds as follows. The next section briefly reviews the Fehr and Schmidt (1999) model and discusses outcome-based preferences more generally. Preferences with variable altruism are introduced in Section 2, where they are used to explain outcomes in ultimatum games. Because agents are assumed to be heterogeneous, the ultimatum game is a signaling game where proposers signal their altruism. As is common with signaling games there are many equilibria though I focus on the one where all possible offers, including the 50-50 division of the pie, are observed. Section 3 is devoted to the dictator game and Section

[6]See Parker *et al.* (2002) for a discussion.

[7]This is not entirely surprising given Kahneman and Tversky's (1979 p. 268) finding that most people prefer \$3000 for sure to a gamble that pays \$4000 with a probability of .8 and zero otherwise while an even larger majority prefers losing \$4000 with probability .8 to losing \$3000 for sure. The former indicates risk aversion while the latter indicates risk-loving in that people choose a prospect that involves a higher variance and a lower expected return.

4 concludes.

1 Outcome-based Preferences

In this section, individuals care both about their own resources and about the way these resources compare with those received by the other player. I focus in particular on the Fehr and Schmidt (1999) model where player i's utility is

$$x_i - \alpha \max(0, x_j - x_i) - \beta \max(0, x_i - x_j) \qquad j \neq i. \tag{1}$$

In this expression, x_i is the level of resources in the hands of agent i, while α and β are positive parameters. People with this utility function wish to hurt those whose income exceeds their own, while they are altruistic towards those whose income is below their own. In an attempt to explain ultimatum outcomes without "fairness," Burnell, Evans, and Yao (1999) use these preferences with $\beta = 0$.

Suppose that the resources available to be split in the ultimatum game equal A. If the responder accepts the proposer's offer of y, she gets y while the proposer keeps $A - y$. Under the assumption that the players' resources outside the experimental setting are either the same or are irrelevant for the utility function (1), one can substitute the payoffs in the experiment for the x's in this function. This implies that responders only accept offers in which $y > \alpha A/(1 + 2\alpha)$. This gives an incentive to proposers to raise their offers above zero. An obvious question, however, is whether this incentive is sufficient to rationalize offers of $A/2$.

Suppose first that α is common knowledge and that $\beta = 0$. Since this implies that the proposer is selfish, he should make the minimum acceptable offer $\alpha A/(1 + 2\alpha)$. This is strictly below $A/2$ unless α is infinite. An infinite α is absurd however, since it implies that individuals are willing to pay any price to achieve a one dollar reduction in the income of a single individual who has more than they do. This leads Fehr and Schmidt (1999) to suppose that α is finite and to use β to rationalize even splits.

For any $\beta < 1/2$, proposers prefer a dollar in their pocket to a dollar in the pocket of

the responder. This means that increasing β above zero has no effect on equilibrium offers until $\beta = 1/2$. When $\beta = 1/2$ the proposer does not care how A is divided so that even splits are possible, though there is no particular reason to suppose that they would emerge in equilibrium. Thus, Fehr and Schmidt (1999) rationalize even splits by supposing that $\beta > 1/2$ so that individuals strictly prefer a dollar in the pocket of someone who has a lower income than themselves to a dollar in their own pocket. Such a high level of altruism seems unreasonable.

Their model has the further disadvantage that individuals with this degree of altruism would also offer even splits in the dictator game. Thus, Fehr and Schmidt (1999) predicts an equal proportion of even split offers in the ultimatum and dictator games, which is contrary to observation.[8]

An alternative method for explaining the outcome of ultimatum games with a utility function of the form of (1) is to follow Costa-Gomes and Zauner (2001) and consider the possibility that $\alpha = \beta < 0$ so that both proposer and responder dislike having the other player receive income. This leads responders to reject ungenerous offers and thereby leads proposers to be generous. Costa-Gomes and Zauner (2001) estimate that the degree of spite needed to explain the results of ultimatum experiments in the United States is quite large, with $\alpha = -.54$. This parameter implies that people ought to be willing to pay 54 cents to cause one dollar of damage to others. This is inconsistent with the willingness of most people to give up small amounts of money if this helps others sufficiently.

The related model of Bolton and Ockenfels (2000) uses a similar rationale for explaining ultimatum outcomes. It differs in explicitly supposing that responders are heterogeneous and this has the obvious advantage of rationalizing the variability in the behavior of responders.

[8]Interestingly, Fehr and Schmidt (1999) contains a discussion of the implications of their setup for dictator games. They argue that utility functions that are curved instead of piecewise linear like (1) would yield offers in dictator games that are strictly between 0 and $A/2$. However, curvature in the utility function can ensure only that people who do not make offers of $A/2$ in ultimatum games make interior offers in dictator games. The right degree of curvature implies that the proposer has increased marginal utility from a dollar in the responder's pocket as offers diminish in size, and thereby guarantees that the responder gets more than zero. Interior offers are indeed observed in dictator experiments. However, with this curvature or without, the fraction of proposers that offer an equal split in the ultimatum game should be the same as the fraction that does so in the dictator game, and this is manifestly counterfactual.

Rather than using the piecewise linear utility function in (1), they consider a differentiable one with diminishing marginal utility of own income. In their model, social preferences are captured by supposing that an individual's utility is concave in the share of total income that he receives with this concave function having a maximum when the income of the two agents is the same.

When an offer is sufficiently uneven, the responder prefers to reject it because this ensures that she has the same income as the proposer. This implies that each responder has a minimum level of y, call it r, such that offers smaller than r are rejected. The parameter r can be interpreted much like α in (1). Consider a responder that is willing to give up r to eliminate the difference in income of $A - 2r$ between proposer and responder. Since the marginal utility for own income is diminishing and the utility due to income shares is concave, such an individual is locally willing to pay at least $\frac{r}{A-2r}$ for each dollar by which she reduces the income difference between the proposer and herself.

Suppose that the distribution of r's among responders can be described by the distribution function $G(r)$. Suppose further that, if the proposal is accepted, the proposer's utility is $A - y + \lambda y$ so that the proposer is altruistic towards the responder with altruism parameter λ. Then, the proposer's utility from making an offer $y \leq A/2$ is

$$G\left(y\right)\Big(A - y(1-\lambda)\Big). \tag{2}$$

An even split then satisfies the first order conditions for a maximum if

$$-G(A/2)(1-\lambda) + G'(A/2)(1+\lambda)\frac{A}{2} = 0. \tag{3}$$

For $\lambda < 1$, which is the reasonable case, this still requires that $G'(A/2) > 0$ so that there is positive density of individuals who are willing to give up $A/2$ for a negligible reduction in inequality (so that they are locally willing to give up an unbounded amount for a reduction by one dollar of the proposer's income). If one is willing to suppose that G' is constant near $A/2$ (which is valid to first order) one can also use this equation to determine the fraction of people whose r must exceed any given threshold $\bar{r}$. This proportion is equal to

$$P(\bar{r}) = \left(\frac{A}{2} - \bar{r}\right)G' = \left(\frac{A}{2} - \bar{r}\right)\frac{2(1-\lambda)}{A(1+\lambda)} = \frac{(1-\lambda)(1-2\bar{r}/A)}{(1+\lambda)}, \tag{4}$$

where the second equality is obtained from (3), taking into account that $G(A/2) = 1$, while the third equality is obtained by simplifying the earlier expression.

Now consider the responders whose r exceeds $2/3$ of $A/2$ so that they are willing to give up $A/3$ to reduce the proposer's income by $2A/3$. Locally, these individuals are willing to pay at least one dollar to ensure that someone richer than themselves loses two dollars, which seems like a degree of mean spiritedness that would lead to a great deal of lawlessness. Using $\bar{r} = A/3$ in (4) implies that $P(\bar{r}) = 1/3$ so that one third of responders must have a higher r for a selfish person to offer an even split. Even if one only requires people with $\lambda = .5$, who are quite altruistic, to offer an even split, it is still necessary for $1/9$ of responders to have an r larger than $A/3$.

As discussed above, Charness and Rabin (2002) do not find their participants willing to incur even much lower costs to hurt people whose incomes are higher. It is important to stress that I am not suggesting that it is possible to explain rejections of ultimatum offers with smaller degrees of ill-will towards proposers. The difficulty is with modelling this ill-will as arising from inequity.

Below, ill-will is allowed to arise endogenously as a reaction to low offers. It is worth stressing, however, that simply stating that an agent's utility depends on other's preferences is not enough. The Levine (1998) model does this, but still requires a large degree of "background" ill-will. That is, it requires that many individuals be quite negatively predisposed towards others even without provocation. The variants of the Rabin (1993) model proposed by Dickinson (2000) and Falk and Fischbacher (2006) do allow for sufficient negative reactions to low offers to induce even splits, but only if reciprocity becomes the only thing that individuals care about (so that agents do not care at all about their own material payoffs). The difficulty with these models seems to be that, like the Levine (1998) model, the functional forms they consider do not allow responders to be sufficiently influenced by low offers. As a result, they seem to require that individuals act at all times as if their utility functions were far from standard.

2 Anger at Insufficient Altruism

Agent i's material payoffs are still assumed to depend only on his resources x_i. However, his utility function W_i is given by

$$W_i = E(x_i + [\lambda^i - \xi(\hat{\lambda}^j, \bar{\lambda}^i)]x_j)^\gamma, \tag{5}$$

where E is the expectation operator and γ represents his attitude towards risk. The expression in square brackets is the individual's altruism. This is positive and equal to λ_i when the individual is not angry, whereas it is negative and equal to $\lambda_i - \bar{\xi}$ when the individual has been angered by the insufficiency of j's altruism. The variable $\bar{\lambda}^i$ represents agent i's anger threshold; it is the minimum level of altruism that he requires of those he interacts with. While individual i knows that different people have different levels of altruism, and such heterogeneity plays an important role below, individual i is assumed to give people the benefit of the doubt. This means that the function ξ is equal to zero unless individual j reveals through her actions that her own altruism parameter is less than $\bar{\lambda}^i$. If she does so, the function ξ takes the value $\bar{\xi}$. One can thus think of individual i as acting as a classical statistician who has a null hypothesis that people's altruism parameter is at least as large as $\bar{\lambda}^i$. If a person acts so that i is able to reject this hypothesis, individual i gains ill-will towards this person.

Since all the parameters of the ultimatum game are known, individuals' actions in this game can be imagined to depend only on their altruism parameter so the statistical testing language is not strictly needed. It is enough to suppose that any demonstration by agent j that his altruism parameter $\hat{\lambda}^j$ is smaller than $\bar{\lambda}^i$ triggers this change in the attitudes of i towards j. More generally, though, people are uncertain not only about the altruism of those they interact with but also about other important aspects of their preferences and constraints.[9] For this case, the statistical testing approach is particularly convenient because it allows agent i to take expectations over all the aspects of individual j about which he is uninformed. If individual j takes an action such that only a fraction smaller than z of

[9] An example of this provided in Section 3 below.

individuals with altruism parameter $\bar{\lambda}^i$ would take an equally ungenerous action, individual i can reject the hypothesis that j's altruism parameter is as a high as $\bar{\lambda}^i$ at the significance level given by z. Individual i is then supposed to change his behavior.

For this general case, $\hat{\lambda}^j$ in (5) represents all the information that i has about j's altruism. The function ξ then takes the value $\bar{\xi}$ if this information lets i reject the hypothesis that λ^j is greater than or equal to $\bar{\lambda}^i$ while it equals zero otherwise. In the special case where actions depend only on altruism, $\hat{\lambda}^j$ is the highest possible value of λ^j that is consistent with the action taken by j and the ξ function equals $\bar{\xi}$ only if $\hat{\lambda}^j < \bar{\lambda}^i$.

While seemingly complicated, the utility function (5) has two attractive properties. The first is that it allows for large changes in attitude in response to small changes in a person's environment. This fits with the observation that people's response to what they perceive to be ill treatment sometimes moves almost seamlessly from passive acceptance to violent outburst, as in the case of road rage. In the context of industrial relations, it is not uncommon for worker unhappiness to erupt in sudden strikes. The second is that agents rarely need to refine their estimate of other people's altruism. Most of the time, they can base their actions on their standard preferences because it is "obvious" that there is no information to suggest that someone is egregiously selfish.

The utility function (5) bears a close connection to the neurological findings of Singer *et al.* (2006). These findings build on several papers (see, for example, Morrison *et al.* 2004) that show activation in pain-related areas of the brain not only when a painful impulse is applied to a subject but also when a subject sees a painful stimulus being applied to someone else. This response, which involves "mirror-neurons," can be thought of as being the neurological analogue of empathy, since individual i is experiencing vicariously the material payoffs of individual j. Singer *et al.* (2006) show that this vicarious response depends on the extent to which j has acted generously in front of i. The brains of subjects that see a confederate making a low offer in a variant of a dictator game have a smaller empathetic response when this confederate experiences an electric shock than the brains of subjects who see the confederate making a generous offer. At the level of the brain, this corresponds to a

change in the subject's altruism parameter.

These neurological observations are closely mirrored by the behavioral observations of Fehr and Fischbacher (2004). They show that agents are willing to incur some costs if they can thereby reduce the payoffs of individuals that have made low offers in a dictator game that these individuals were playing with *third parties*.[10]

This utility function can be simplified further in the ultimatum game because the proposer gets no information about the responder so that he accords her the benefit of the doubt. Using p and r to denote the proposer and responder respectively, the proposer's utility is thus

$$W_p = E(x_p + \lambda^p x_r)^\gamma. \tag{6}$$

The expectations operator is present here because the $\bar{\lambda}^i$ of the responder dictates whether the offer is accepted and, when responders are heterogeneous, this parameter is unknown to the proposer. For the responder, by contrast, the outcome is a deterministic function of whether he accepts or rejects the offer. This means that, as long as $\gamma > 0$, the actions that maximize the utility function (5) of the responder also maximize this utility function when γ is set to one. For simplicity, I thus let the responder maximize

$$W_r = x_r + (\lambda^r - \xi(\hat{\lambda}^p, \bar{\lambda}))x_p, \tag{7}$$

where $\bar{\lambda}$ is used rather than $\bar{\lambda}^r$ to simplify the notation. Let $\bar{\lambda}$ be randomly distributed among potential responders with a distribution function $H(\bar{\lambda})$ so that different people demand a different level of benevolence from those they interact with in this game. The support for this distribution is the set $[\lambda_L, \lambda_H]$.

If $\bar{\xi}$ has an arbitrarily large value, responders are willing to incur arbitrarily large costs to punish proposers whose λ^p is below $\bar{\lambda}$. In practice, most people's willingness to incur costs

[10]In Fehr and Fischbacher (2004), the punishing players can induce a loss of \$3 to a "bad" player for each dollar they give up themselves. Fehr and Fischbacher (2004) claim that the Fehr and Schmidt (1999) model can explain these punishments because these punishments reduce the difference between the payoff to the person who acted greedily in the dictator game and the person who is able to punish. If this explanation were correct, these punishments would disappear if the person who can punish were given a larger endowment in the experimental setting. This seems unlikely.

to punish those that are insufficiently altruistic appears to be bounded. This suggests that $\bar{\xi}$ should be bounded above. A simple condition on $\bar{\xi}$ which has the additional advantage of ensuring that responders are willing to accept offers of even splits regardless of the λ^p that they imply is $(\bar{\xi} - \lambda^r) = 1$. This implies that the utility of responders who feel mistreated is a function of $(x_r - x_p)$ so that responders are not willing to pay more than a dollar to cause proposers one dollar worth of damage. This means that responders are strictly better off accepting offers for which $x_r > x_p$ while they are indifferent to accepting or rejecting offers of even splits. Imposing this condition while supposing that responders accept offers in the case of indifference thus rationalizes the general acceptance of even splits, and I rely on this acceptance below. [11]

The last parameter that deserves discussion is γ. It is important to stress that the conclusions of this paper do not depend on whether this parameter is interpreted as corresponding to risk attitudes or whether it captures some other aspect of preferences for small gambles including the utility of gambling discussed in Conlisk (1993). Conlisk (1993) discusses several experiments that mimic people's willingness to pay for the use of slot machines. On the other hand, it must be said that the experimental literature that lets subjects choose among small stake gambles does not consistently report risk-loving (or risk-averse) choices.[12] Battalio, Kagel and Jiranyakul (1990) display a large battery of such experiments, many of which bear some similarity to the problem confronting proposers in the ultimatum game. In their Table 3 (p. 34), they report 6 experiments that ask subjects to choose between a known gain and a lottery with the same expected value and two possible positive outcomes. In all six of these experiments, a majority of subjects chose the risky outcome. By contrast,

[11] While deriving the condition $(\bar{\xi} - \lambda^r) = 1$ from first principles is beyond the scope of this paper, equal splits have obvious advantages from a risk spreading perspective. This means that these values of $\bar{\xi}$ may have desirable evolutionary properties. People who agree to disagree on the minimal level of altruism might ultimately agree that it is socially costly to adopt different levels of $\bar{\xi}$. It would also be interesting to develop models in which even splits are accepted for other psychological reasons, including the desire of responders to act in ways that they themselves regard as reasonable.

[12] Holt and Laury (2002) confront subjects with two gambles, both of which are risky. In their study, most subjects made decisions consistent with risk aversion. By contrast, the vast majority of Kachelmeier and Sehata's (1992) subjects acted in a risk loving manner when they were asked to name the payment they would require to give up the right to earn a high prize with low probability.

their Table 2 (p. 33) reports some experiments with the same structure where a majority prefers the safe choice. One difference between the two sets of results is that risk aversion seems more common when the probability of the better of the two random outcomes is higher (so that the lottery is less similar to the distribution of outcomes in gaming tables and slot machines). These findings obviously do not pin down γ, but they do suggest that values of γ somewhat larger than one should not be discarded as implausible for this type of setting.

Let the offer y of the proposer take values in a discrete grid and denote its equidistant values by $y_0, y_1, \ldots, y_n$, where $y_0 = 0$ and $y_n = A/2$.[13] Because this is a signaling model where the offer y signals the type λ^p, there are numerous equilibria, which are supported by different beliefs regarding off-the-equilibrium-path behavior. I focus on equilibria with two properties. The first is that the highest offer is equal to $A/2$. Not only is this essential for fitting the experimental results, but it also has the attractive feature of being a conservative (or safe) offer for those proposers whose altruism parameter λ^p is greater than or equal to λ_H, so that their altruism is as large as that demanded by any responder. For these proposers, an offer of $A/2$ is optimal if these proposers are as pessimistic as possible about the beliefs of responders, *i.e.* if the proposers believe that responders associate lower offers with substantially lower values of λ so that they accept lower offers with lower probability.

I also require that each possible offer be made by some proposer in equilibrium. This, again, fits well with experimental findings. It also means that the equilibrium outcome is not caused by "unreasonable" off-the-equilibrium beliefs; responders' beliefs are reasonable because they correspond to actions that proposers with different λ^p's actually take. This has the benefit of ensuring that the conservatism I impose on proposers with $\lambda^p = \lambda_H$ is reasonable, since offers below $A/2$ do, in fact, lead to lower probabilities of acceptance.

Given that the set of offers is discrete while λ^p can take a continuum of values, proposers with different λ^p's must make the same offer. Let $\tilde{\lambda}_i$ represent the set of λ^p's that make offer y_i, and let λ_i be the supremum of $\tilde{\lambda}_i$. Responders who receive the offer y_i cannot dismiss

[13] I have also studied the case where y can take a continuum of values. While the mass of proposers who offer even splits is more difficult to study in this case, the continuum formulation has the attractive feature that the resulting objective function is very similar to (2).

the possibility that this offer was made by someone whose altruism parameter is equal to λ_i. The model then implies that responders for whom $\bar{\lambda} \leq \lambda_i$ give the proposer the benefit of the doubt and accept the offer. The offer y_i is thus accepted with probability $H(\lambda_i)$. Therefore, an equilibrium is a collection of disjoint sets $\tilde{\lambda}_i$ such that $\forall \lambda \in \tilde{\lambda}_i, \forall j \neq i$,

$$H(\lambda_i)(A - y_i(1-\lambda))^\gamma \geq H(\lambda_j)(A - y_j(1-\lambda))^\gamma. \tag{8}$$

Simple inspection of the proposer's objective function makes it clear that, if the probability of acceptance H is increasing in the proposer's offer y, proposers with higher values of λ prefer higher values of y. By the same token, if proposers with higher λ^p tend to make higher offers, higher offers lead fewer responders to reject the hypothesis that $\lambda^p \geq \bar{\lambda}$ so that offers are more likely to be accepted. The model thus satisfies Athey's (2001) single-crossing property and her analysis then implies that a weakly monotone equilibrium in pure strategies exists. In the equilibria I compute, this monotonicity is strong so that whenever $y_i < y_j$, $\lambda_i < \lambda_j$ and $H(\lambda_i) < H(\lambda_j)$.

Consider an equilibrium where some proposers offer y_i while others offer y_{i+1}. Such an equilibrium requires that proposers with $\lambda^p = \lambda_i$ for $i < n$ must be indifferent between y_i and y_{i+1}. To see this note that, since this proposer chooses y_i, he cannot prefer y_{i+1}. And if he strictly preferred y_i, a proposer with a slightly higher λ^p would prefer y_i as well, thereby violating the definition of λ_i. Thus,

$$H(\lambda_i)(A - y_i(1-\lambda_i))^\gamma = H(\lambda_{i+1})(A - y_{i+1}(1-\lambda_i))^\gamma. \tag{9}$$

Once one fixes the probabilities $H(\lambda_i)$ and $H(\lambda_{i+1})$, the right hand side of (9) rises faster with λ_i than the left hand side. This means that, for given cutoffs λ_i and λ_{i+1}, all individuals with altruism parameters larger than λ_i prefer the higher offer.

At an equilibrium where all offers are made, the equilibrium cutoffs for the altruism parameter λ^p must satisfy the difference equation (9) for all $0 \leq i < n$. Thus, $\tilde{\lambda}_i = (\lambda_{i-1}, \lambda_i]$ for $i > 0$ while $\tilde{\lambda}_0 = [\lambda_L, \lambda_0]$. For the highest offer to be $A/2$, the relevant solution to this difference equation must have the boundary condition $\lambda_n = \lambda_H$. Such an equilibrium exists

as long as the λ_0 one obtains from applying (9) satisfies $H(\lambda_0) \geq 0$. If this holds, no proposer whose λ^p is in $\tilde{\lambda}_i$ wishes to make any offer other than y_i. If it does not hold, the value of λ_0 one obtains does not make sense so that the other values of λ_i do not either.

Fortunately, it is immediately apparent that the solution of (9) with $\lambda_n = \lambda_H$ does indeed satisfy $H(\lambda_0) \geq 0$. To see this, note first that $(A - y_{i+1}(1 - \lambda_i))^\gamma$ is positive if $0 \leq y_i \leq A/2$ and $\lambda_i < 1$, as I assume. This implies that whenever the right hand side of (9) is positive for a given i, $H(\lambda_i)$ is positive, implying that the right hand side of (9) is positive for $i - 1$. Since the right hand side of (9) is positive for $i = n - 1$, it follows that $H(\lambda_0)$ is positive as well.

One interesting feature of this equilibrium is that the actions of responders, and thus the offers made by a proposer with a given λ^p, are independent of the distribution of λ^p itself.[14] They depend only on the distribution of the $\bar{\lambda}$'s, the levels of altruism that people regard as minimally acceptable.

Offers made by proposers with different λ^p's are illustrated, together with their probability of being accepted, in Figure 1. This figure shows solutions to the difference equation in (9) for the base case where $\gamma = 1$ and where the distribution $H(\bar{\lambda})$ is uniform between 0 and 0.2. It displays equilibria both for the case where the grid size equals one tenth of the pie to be distributed (so that in a \$10 experiment, the proposers make integer offers) and for a grid size equal to 5E-5 times the size of the pie, which is obviously much smaller. The figure fits a smooth line through the offers that are made in the two cases. Thus, it should be interpreted as saying that the probability of acceptance of offers that are common in both situations (such as 0.4 of the pie) are extremely similar. Also, among responders that accept a particular offer (say 0.4 of the pie) the one with the highest $\bar{\lambda}$ has a very similar $\bar{\lambda}$ in the two cases. In the case of offers of 0.4 of the pie, this highest $\bar{\lambda}$ equals 0.1753 in the case of

[14] In Levine (1998), by contrast, the distribution of λ's does matter for the probability that an offer will be accepted by responders. In his model, the probability of accepting an offer depends on the altruism of responders and he supposes that the distribution of this altruism parameter is the same as the distribution of the altruism parameter among proposers. By contrast, the altruism of responders plays a more muted role in my analysis because this altruism is swamped by the anger that attends rejection of the hypothesis that a proposer's λ^p is at least equal to $\bar{\lambda}$.

the coarse grid and 0.1760 in the case of the fine one. On the other hand, the lowest $\bar{\lambda}$ that ends up accepting only a particular offer is smaller in the case of a coarse grid, because the fine grid ensures that there are smaller offers that are still acceptable to this more tolerant responder.

In this rational expectations equilibrium, the range of $\bar{\lambda}$'s of responders that find a particular offer y_i minimally acceptable corresponds to the range of λ's of proposers that make this particular offer. This means, not surprisingly, that the range of proposers that make a particular integer offer is larger when the offer grids is coarser and there is smaller menu of offers available. By the same token, even splits ought to become less common when proposers have a finer grid to choose from.

This raises the question of whether this model can account for the fact that between 28 and 50 percent of proposers offer even splits. This turns out to be crucially dependent on how the distribution of λ^p's compares to the distribution of $\bar{\lambda}$'s. If the two coincide, the parameters underlying Figure 1 cannot account for the high fraction of such offers, even with the coarse grid based on integer offers. With a uniform distribution between 0 and 0.2, only about 12 percent of responders have $\bar{\lambda}$'s between 0.175 and 0.2, so that this is the proportion of responders that accept only an even split. If λ^p had the same distribution, only about 12 percent of proposers would offer even splits. On the other hand, it seems more reasonable to suppose that people with relatively high levels of altruism have $\bar{\lambda}$'s below their own λ.[15] One would then expect many more proposers to have λ^p's above 0.175 so that the proportion of even splits would exceed 12 percent , possibly by a large margin.

Figure 2 maintains the coarse grid but varies some aspects of the distribution of $\bar{\lambda}$. It compares a uniform distribution between 0 and 0.2 to a uniform distribution between 0 and 0.1. It also considers a density whose range goes from 0 to 0.1 but where low values of $\bar{\lambda}$ are relatively more common. By letting the probability density function of $\bar{\lambda}$ equal $35 - 500\bar{\lambda}$,

[15] Price, Cosmides, and Tooby (2002) show that people who are more willing to contribute to a public good are also more likely to want to punish non-contributors. While this suggests that individuals with a higher λ tend to have a higher $\bar{\lambda}$, it does not bear on the question of whether the distributions of λ and of $\bar{\lambda}$ are identical.

the density of $\bar{\lambda}$ at zero becomes 25 times larger than the density at 0.1. Moving from the uniform distribution on the range $[0, .1]$ to this pdf has a similar effect to moving from the uniform distribution in the range $[0, .2]$ to the uniform distribution in the range $[0, .1]$. In both cases, the probability of acceptance falls somewhat more rapidly as offers decline, though the differences are modest.

Some intuition for these differences can be developed by considering a linearized version of the difference equation (9), where the linearization is carried out around the point λ_i, y_i. After dividing through by $(A - y_i(1 - \lambda_i))^{\gamma-1}$ this gives

$$h(\lambda_i)(A - y_i(1 - \lambda_i))(\lambda_{i-1} - \lambda_i) - \gamma H(\lambda_i)(1 - \lambda_i)(y_{i-1} - y_i) \approx 0, \tag{10}$$

where $h(x)$ is the density of H at x. The second term is the increase in the proposer's gain from offering a lower y, while holding constant the probability of acceptance H. The first term, meanwhile, is approximately equal to the proposer's loss from the reduction in the probability that his offer will be accepted, where $(A - y_i(1 - \lambda_i))$ is the amount he loses and $h(\lambda_i)(\lambda_{i-1} - \lambda_i)$ is approximately the reduction in the acceptance probability. This equation can be rearranged so that the change in the probability of acceptance that results from offering y_{i-1} rather than y_i is

$$h(\lambda_i)(\lambda_{i-1} - \lambda_i) \approx \gamma \frac{H(\lambda_i)(1 - \lambda_i)}{A - y_i(1 - \lambda_i)}. \tag{11}$$

Since $\lambda_n = \lambda_H$ and $H(\lambda_H) = 1$, (11) implies that a higher value of λ_H leads to a slower decline in the probability of acceptance as offers are reduced from $A/2$. There are essentially two reasons for this. The first is that a higher λ_i lowers the extent to which a lower y raises utility, because the individual empathizes more with the responder's loss. This lowers the second terms of (10) and implies that a smaller reduction in the probability of acceptance is needed to keep the proposer indifferent with respect to a reduction in y. The second reason is that a higher λ implies that the loss from having a given fall in the probability of acceptance (which is captured by the first term of (10)) is larger, precisely because the proposer is also losing the vicarious benefit he draws from the resources received by the responder. This also

means that a smaller decline in the probability of acceptance is warranted to compensate for the gain that results from a given reduction in y.

These two effects explain the differences in acceptance probabilities displayed in Figure 2. When moving from having $\bar{\lambda}$ distributed uniformly on $[0, .2]$ to having it distributed uniformly on $[0, .1]$, λ_H falls, and these two effects lead the probability of acceptance to fall more rapidly with y. Now consider the non-uniform distribution in Figure 2. Equation (11) implies that the reduction in the probability of acceptance as y is reduced from $A/2$ depends only on the value of λ_H and is thus the same in the case of the uniform distribution over the range $[0, .1]$ as in the case of this non-uniform distribution. However, the density of $\bar{\lambda}$ at λ_H is much smaller when the pdf of $\bar{\lambda}$ is $35 - 500\bar{\lambda}$. This implies that, for a given reduction in the probability of acceptance, the distance between λ_n and λ_{n-1} must be greater and λ_{n-1} must be lower. The earlier argument then implies that the reduction in the probability of acceptance between y_{n-1} and y_{n-2} must be larger in the case of this non-uniform density.[16]

Even though this argument shows that changes in the distribution of $\bar{\lambda}$ can change the speed at which the probability of acceptance falls with y, the parameters I have considered so far do not allow the probability to fall by enough to explain the evidence. Combining the results of various studies, Harrison and McCabe (1996) compute that the probability that offers of zero will be accepted equals 0.33. They also analyze data originally collected for Carter and Irons (1991) on how responders would react to various potential offers and show that only 15 percent of these responders would accept such an offer. By contrast, Figures 1 and 2 show that the probability of accepting offers as low as zero always exceeds 0.5.

For $\gamma = 1$, this turns out to be guaranteed if λ_L, the minimum level of altruism that anyone finds acceptable, is nonnegative. The reason is as follows. Any risk neutral selfish responder strictly prefers to offer an even split if this is accepted with probability one to having a zero offer rejected with a probability larger than 0.5. This means that anyone who makes a zero offer that has a probability of being accepted lower than 0.5 must be spiteful,

[16]The analysis above was carried out for $H(\lambda_i) = 1$ and $H(\lambda_{n-1})$ is a bit smaller. But, to first order, this effect is negligible.

i.e. must have a $\lambda^p < 0$. Since this is unacceptable to all responders, such an offer could not be accepted. This, in turn, contradicts the possibility that an offer of zero would be made if it had a probability less than 0.5 of being accepted.

This argument extends naturally to any offer that leaves the proposer with expected earnings lower than $A/2$. Since a proposer can guarantee himself this level of earnings by offering $A/2$, a proposer with $\gamma = 1$ would make a lower offer that led to lower expected earnings only if he were spiteful, and this would make such an offer unacceptable. If, by contrast, proposers liked small gambles and responders knew this, such offers would become acceptable at least to some responders.

To see this, Figure 3 shows acceptance probabilities for γ equal to both 2 and 3, where the distribution of H remains equal to the uniform distribution on $[0, .2]$. The figure also displays the acceptance probabilities computed by Harrison and McCabe (1996) from a variety of previous studies, as well as those they obtained analyzing the data of Carter and Irons (1991). The figure shows that most of these data are consistent with the model with $\gamma = 2$. The exception is the probability of acceptance of zero offers obtained from the Carter and Irons (1991) data, which fits this model only when $\gamma = 3$. Leaving aside this observation, it appears that a moderate taste for small gambles can explain why proposers whose λ^p is smaller than λ_H offer y's lower than $A/2$ even though the acceptance probabilities that result from such offers are significantly lower than those that emerge in equilibrium when $\gamma = 1$.

3 The Dictator Game

If proposers in the dictator game could not be penalized at all for making small offers, and if their altruism parameter were smaller than one, their nonsatiation would imply that they would offer zero. This suggests that proposers who can be sure that their offers will remain anonymous ought to offer less than those who fear detection of their offers by others. This effect has been demonstrated experimentally. Hoffman *et al.* (1994) conducted a dictator experiment in which they increased the proposers' anonymity vis-a-vis the experimenter

and found that this reduced the offers made by proposers.[17] Similarly, Burnham (2003) showed that letting proposers know that responders would see pictures of them considerably increased their offers. Anonymity for proposers is difficult to assure for two reason. First, the proposer may fear that the experimenter will let others know the size of his offer.[18] Second, the proposer may not trust himself not to reveal his offer to others. Many studies have shown that lying can lead to telltale signs that can be detected. The policy of telling the truth may thus be individually rational for many people.

In this section, I therefore suppose that the proposer in the dictator game feels that there is a positive probability π that his offer will be discovered by someone whom I call the responder. If the responder concludes that the dictator's altruism is below $\bar{\lambda}$, the dictator suffers a loss V. For this analysis, it is not important that the responder be the recipient of the offer. The results of Fehr and Fischbacher (2004) suggest instead that a variety of individuals are prepared to punish people whom they regard as having been insufficiently generous in the dictator game.

This mechanism of punishment may well have heterogenous effects across people. Some people may, for example, be better at concealing their actions from others so that their π is lower. Along the same lines, some people may have a bigger preference for telling others about their experiences during the experiment. They might also differ in the reaction they expect others to have when they find out that they made a low offer. This means that proposers' offers do not depend only on the proposers' altruism parameters but also on their beliefs concerning π and V. Responders, on the other hand, want to punish only those proposers whose altruism they regard as insufficient. If responders do not know π or V, they can use the statistical approach in (5) and punish only those proposers whose actions are sufficiently unlikely to be caused by a proposer with the proper $\bar{\lambda}$. To simplify this section, I suppose instead that responders actually know the proposer's π and V, so that the offer

[17] Their clever experimental design reduced the evidence left behind by low offers because the offer's only physical record was the currency received (via an envelope) by the responder (as well as the currency left in the proposer's pocket).

[18] In Ben-Ner *et al.* (2004), this is done without warning the proposers beforehand.

contains information only about the proposer's altruism parameter.

This assumption allows the earlier analysis to apply almost directly. In particular, for each offer y_i, let $\tilde{\lambda}_i(V,\pi)$ represent the set of λ^p's that make offer y_i given their probability of detection π and their cost of punishment V. Further, let $\lambda_i(V,\pi)$ once again be the supremum of $\tilde{\lambda}_i(V,\pi)$. Responders who know V and π while observing an offer of y_i are willing to treat the proposer as if his altruism parameter were equal to λ_i. This means that the probability that a responder would refrain from punishing such a proposer is $1-H(\lambda_i)$. The sets $\tilde{\lambda}_i$ are thus an equilibrium if $\forall\lambda\in\tilde{\lambda}_i, \forall j\neq i$,

$$(A-y_i(1-\lambda))^\gamma-\pi(1-H(\lambda_i))V\geq(A-y_j(1-\lambda))^\gamma-\pi(1-H(\lambda_j))V. \tag{12}$$

As before, higher values of λ make higher values of y_i more valuable to proposers if responders are more likely to punish proposers whose y is lower. And similarly, responders who know that higher offers are made by individuals with higher λ^p's are more likely to punish proposers whose offers are low. There is thus a separating equilibrium where higher offers are made by proposers with higher λ^p's for given V and π. One immediate consequence of (12) is that, for given $\tilde{\lambda}$'s, the only individual-specific parameter that affects the behavior of the proposer is the product πV.

Suppose as before that some proposers with a given πV offer y_i while others offer y_{i+1}. Then, proposers with this πV for whom $\lambda^p=\lambda_i$ must be indifferent between y_i and y_{i+1}. If they liked y_i less, (12) would be violated and, if they liked it more, proposers with $\lambda^p<\lambda_i$ would choose y_{i+1}, thereby contradicting the definition of λ_i. Therefore

$$(A-y_i(1-\lambda_i))^\gamma-\pi V(1-H(\lambda_i))=(A-y_{i+1}(1-\lambda_i))^\gamma-\pi V(1-H(\lambda_{i+1})), \tag{13}$$

and the sets $\tilde{\lambda}_i$ consist of the sets $(\lambda_{i-1},\lambda_i]$. The single crossing property then ensures that proposers prefer offering y_{i+1} to offering y_i if and only if their λ^p exceeds λ_i. Equation (13) takes on a particularly simple form when $\gamma=1$ and H is uniform in the range $[0,\lambda_H]$, so $H(\lambda_i)=\lambda_i/\lambda_H$. The equation can then be rearranged to yield

$$\left(\frac{\pi V}{\lambda_H}-\delta_y\right)\lambda_i=\frac{\pi V}{\lambda_H}\lambda_{i+1}-\delta_y, \tag{14}$$

where δ_y is the size of the offer grid $y_{i+1} - y_i$. Neither equation (13) nor (14) can constitute an equilibrium unless decreases in λ_{i+1} lead to reductions in λ_i. This requires that the gap between grid points δ_y be sufficiently small, and specifically in the case of (14), that it be smaller than $\pi V/\lambda_H$.

The solution of the difference equation (14) requires a boundary condition. To obtain this boundary condition, I once again seek an equilibrium that satisfies two properties. The first is that proposers with $\lambda^p = \lambda_H$ be as pessimistic as possible about the offer that will avoid the punishment V, so that they make the highest possible offer that is consistent with equilibrium. As I demonstrate below, this no longer needs to equal $A/2$. The second is that all offers below the offer made by proposers with $\lambda^p = \lambda_H$ be made in equilibrium. As before, this ensures that the equilibrium is supported by punishments that are actually observed in equilibrium play.

The second of these conditions implies that $\lambda_0 > 0$ so that proposers with sufficiently low altruism offer nothing (*i.e.* propose y_0). Suppose, on the other hand, that the proposers whose altruism parameter equals λ_H offer y^m. Applying equation (14) implies the following link between λ_0 and y_m:

$$\lambda_0 = \left(\frac{\pi V}{\pi V - \lambda_H \delta_y}\right)^m (\lambda_H - 1) + 1. \tag{15}$$

With $\lambda_H < 1$, a higher value of m (or y_m) reduces the value of λ_0 that is consistent with (15). The intuition for this finding is the following. If the proposers with the highest level of altruism make higher offers, people with somewhat lower levels of altruism must make higher offers as well. This is needed to ensure that the people with altruism levels λ_i are indifferent between offering y_i and y_{i+1}. As a result, only lower altruism individuals are left to make offers of zero.

Equation (15) implies that there is a maximum level of m that is consistent with $\lambda_0 > 0$. If

$$\left(\frac{\pi V}{\pi V - \lambda_H \delta_y}\right)^m (\lambda_H - 1) + 1 > 0, \tag{16}$$

an even split is below this maximal value. Otherwise it is lower.

When (16) is satisfied, it is reasonable to suppose that proposers with $\lambda^p = \lambda_H$ do not offer more than an even split even if the maximum possible punishment allows one to sustain even more generous offers. When (16) is violated, the equilibrium with the highest possible value of m such that $\lambda_0 > 0$ is the one that is most conservative for proposers with $\lambda^p = \lambda_H$ among those that satisfy the second property suggested above. It is thus the equilibrium I focus on.

One implication of (15) is that reductions in the expected punishment πV reduce the maximum sustainable value of y_m, and the same is true of reductions in the maximum altruism of proposers λ_H. The intuition for these results is straightforward. Lower values of πV and of λ_H make it less costly for proposers with altruism parameter λ_H to offer zero because the maximum possible loss is πV. Therefore, they make it more difficult to sustain higher offers in equilibrium.

The model thus has a straightforward explanation for the observation that low offers are much more common in dictator games than in ultimatum games. This is that, among proposers with $\lambda^p = \lambda_H$, only those for whom πV is large continue to offer even splits in this game. Proposers with lower values of πV make less generous offers even if their altruism level equals λ_H.

4 Conclusions

This paper has presented a model of preferences that can account for the experimental findings of ultimatum and dictator games without imposing extreme parameter values. It supposes that people feel mildly altruistic towards one another in most circumstances. Their normal altruism is mild enough that they would not transfer a dollar from their pocket to someone with a similar marginal utility of income, though they would transfer resources to people whose marginal utility of income they perceive as being much higher. They would also be willing to give up a dollar if someone else thereby gained substantially more than a dollar, as in the experiments of Charness and Rabin (2002).

The reason why preferences that differ so little from the selfish ones that form the baseline

of economic analysis can fit these experiments is that there is a trigger that leads people to have very different preferences. In particular, people get upset with people who demonstrate extreme selfishness. One way of thinking about this is that the model formalizes the Camerer and Thaler (1995) insight that people get angry at individuals who have "poor manners." Once this reaction has been triggered, people actually enjoy hurting those that they regard as excessively selfish. What makes the ultimatum and dictator games different from more normal economic interactions is that they are good litmus tests for the extent to which people are selfish, because actions in these games signal little else.

By contrast, the moves people make in other economic interactions typically signal also their tastes for different commodity bundles, as opposed to the extent of their altruism. These interactions are thus less prone to trigger anger. To see this, consider two individuals who wish to buy the same good. Suppose first that the price is fixed, that there is only one unit of the good left and that the first manages to purchase the unit before the second. Even if the second covets the good as well, the purchase by the first does not necessarily prove that this individual is ungenerous. Thus, the second individual's disappointment is unlikely to turn to anger.

Alternatively, imagine that the two individuals are bidding for the single unit that they both desire. When the first individual raises his bid this raises the cost to the second of obtaining the unit. Even so, the second is not entitled to see this as purely reflecting the first individual's selfishness. This is so not only because the first individual might desire the good intensely but also because the first individual might be equally altruistic towards the seller as towards the second individual, and such even handedness does not seem as subject to censure. This even handedness would imply that, if the second individual ends up with the good, the vicarious gain to the first individual from the resources gained by the seller equals the vicarious losses he experiences from the reduction in the resources available to the buyer. There is thus no altruistic gain to the first individual from keeping his bid low. This can explain why selling prices end up being close to the reservation price of buyers

in the "market" experiment of Roth *et al.* (1991).[19] In this experiment there are several potential buyers so that the capacity of the subjects to experience anger does not lead to price restraint. Matters are different when there is a single buyer and a single seller. In this case, a seller that posts a high price demonstrates that he is selfish and this can lead to punishment. Hoffman *et al.* (1994) display an ultimatum game with this structure and show that, indeed, sellers end up charging considerably less than the reservation price of the buyer.

Still, unlike the case of the ultimatum game with full information, buyers do not typically know the cost conditions faced by sellers, and this means that a seller's price is a less accurate measure of the seller's altruism than is the proposer's offer in an ultimatum game. Thus, price setting is more similar to the ultimatum game with incomplete information introduced by Mitzkewitz and Nagel (1993). I consider a price-setting situation of this sort in Rotemberg (2005) and show that, under plausible conditions, a single-shot price can be a quite poor signal of the seller's altruism. This would suggest that producers have a great deal of flexibility regarding the prices they charge. However, the ability of producers to change their prices without triggering anger is substantially lessened if buyers do not believe that cost conditions have changed significantly from the time that prices were changed last.

[19]Note also that, as long as the seller's altruism parameter is less than one, he would choose to sell to the buyer that offers the highest price. Choosing this buyer therefore does not prove that the seller is excessively selfish.

References

Athey, S., 2001. Single Crossing Properties and the Existence of pure Strategy Equilibria in Games of Incomplete Information. Econometrica 69, 861-889.

Battalio, R.C., Kagel, J.H., Jiranyakul, K., 1990. Testing Between Alternative Models of Choice under Uncertainty: Some Initial Results. Journal of Risk and Uncertainty 3, 25-50.

Ben-Ner, A., Putterman, L., Kong, F., Magand, D., 2004. Reciprocity in a two-part dictator game. Journal of Economic Behavior & Organization 53, 333-352.

Blount, S., 1995. When social outcomes aren't fair: The effect of causal attributions on preferences. Organizational Behavior and Human Decision Processes 63, 131-144.

Bolton, G.E., Ockenfels, A., 2000. ERC: A Theory of Equity, Reciprocity, and Competition. American Economic Review 90, 166-193.

Burnell, S.J., Evans, L., Yao, S., 1999. The Ultimatum Game: Optimal Strategies without Fairness. Games and Economic Behavior 26, 221-252.

Burnham, T.C., 2003. Engineering Altruism: A Theoretical and Experimental Investigation of Anonymity and Gift Giving. Journal of Economic Behavior & Organization 50, 133-144.

Camerer, C.F., Thaler, R.H., 1995. Ultimatums, Dictators and Manners. Journal of Economic Perspectives 9, 209-219.

Carter, J.R., Irons, M.D., 1991. Are Economists Different, and If So, Why?. Journal of Economic Perspectives 5, 171-77.

Charness, G., Rabin, M., 2002. Understanding Social Preferences with Simple Tests. Quarterly Journal of Economics 117, 817-869.

Choi, J.J., Laibson, D., Madrian B.C., Metrick, A., 2004, For Better or For Worse: Default Effects and 401(k) Savings Behavior, Perspectives on the economics of aging, University of Chicago Press: Chicago, 81-121.

Conlisk, J., 1993. The Utility of Gambling. Journal of Risk and Uncertainty 6, 255-275.

Costas-Gomes, M., Zauner, K.G., 2001. Ultimatum Bargaining Behavior in Israel, Japan, Slovenia and the United States: A Social Utility Analysis. Games and Economic Behavior 34, 238-269.

Dickinson, D.L., 2000. Ultimatum Decision Making: A Test of Reciprocal Kindness. Theory and Decision 48, 151-177.

Falk, A., Fischbacher, U., 2006. A Theory of Reciprocity. Games and Economic Behavior 54, 293-315

Fehr, E., Fischbacher, U., 2004. Third Party Punishment and Social Norms. Evolution and Human Behavior 25, 63-87.

Fehr, E., Schmidt, K.M., 1999. A Theory of Fairness, Competition and Cooperation. Quarterly Journal of Economics 114, 817-868.

Forsythe, R., Horowitz, J.L., Savin N.E., Sefton, M., 1994, Fairness in Simple Bargaining Games. Games and Economic Behavior 6, 347-369.

Güth, W., Schmittberger, R., Schwarze, B., 1982. An Experimental Analysis of Ultimatum Bargaining. Journal of Economic Behavior and Organization 3, 367-388.

Harrison, G.W. McCabe, K.A., 1996. Expectations and Fairness in a Simple Bargaining Experiment. International Journal of Game Theory 25, 303-327.

Henrich J., Boyd, R., Bowles, S., Camerer, C.F., Gintis, H., 2004. Overview and Synthesis. In Henrich J., Boyd, R., Bowles, S., Camerer, C.F., Gintis, H. (Eds), Foundations of Human Sociality, Oxford University Press: Oxford, 8-54.

Hoffman, E., McCabe, K., Shachat, K., Smith, V., 1994, Preferences, Property Rights and Anonymity in Bargaining Games. Games and Economic Behavior 7, 346-380.

Holt, C.A. Laury, S.K., 2002. Risk Aversion and Incentive Effects. American Economic Review 92, 1644-1655.

Kachelmeier, S.J., Sehata, M., 1992. Examining Risk Preferences under High Monetary Incentives; Experimental Evidence from the People's Republic of China. American Economic Review 82, 1120-1141.

Kagel, J.H., Wolfe, K., 2001. Tests of Fairness Models Based on Equity Considerations in a Three-Person Ultimatum Game. Experimental Economics 4, 203-219.

Kahneman, D. Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica 47, 263-292.

Levine, D.K., 1998. Modelling Altruism and Spitefulness in Experiments, Review of Economic Dynamics 1, 593-622.

Loewenstein, G.F., Thompson, L. Bazerman, M.H., 1989. Social utility and decision making in interpersonal contexts. Journal of Personality and Social Psychology 57, 426-441.

Mitzkewitz, M., Nagel, R., 1993. Experimental Results on Ultimatum Games with Incomplete Information. International Journal of Game Theory 22, 171-198.

Morrison, I., Lloyd, D., Di Pellegrino, G., Roberts, N., 2004. Vicarious responses to pain in anterior cingulate cortex: Is empathy a multisensory issue?. Cognitive, Affective and Behavioral Neuroscience 4, 270-278.

Parker, D., Lajunen, T., Summala, H., 2002. Anger and aggression among drivers in three European countries. Accident Analysis and Prevention 34, 229-235.

Pillutla, M.M., Murnighan, J.K., 1996. Unfairness, Anger, and Spite: Emotional Ultimatum Offers. Organizational Behavior and Human Decision Processes 68, 208-224.

Price, Michael E., Cosmides, L., Tooby, J., 2002. Punitive sentiment as an anti-free rider psychological device. Evolution and Human Behavior 23, 203-231.

Rotemberg, J.J., 2005. Customer Anger at Price Increases, Changes in the Frequency of Price Adjustment and Monetary Policy. Journal of Monetary Economics 52, 829-852.

Rabin, M. 1993. Incorporating Fairness into Game Theory and Economics. American Economic Review 83, 1281-1302.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and Market Behavior in Jerusalem, Liubljana, Pittsburgh and Tokyo: an Experimental Study. American Economic Review 81, 1068-1095.

Singer, T., Seymour, B., O'Doherty, J.P. *et al.* 2006. Empathic neural responses are modulated by the perceived fairness of others. Nature 439, 466-469.

Figure 1: Probability of Acceptance: Variations in Grid Size

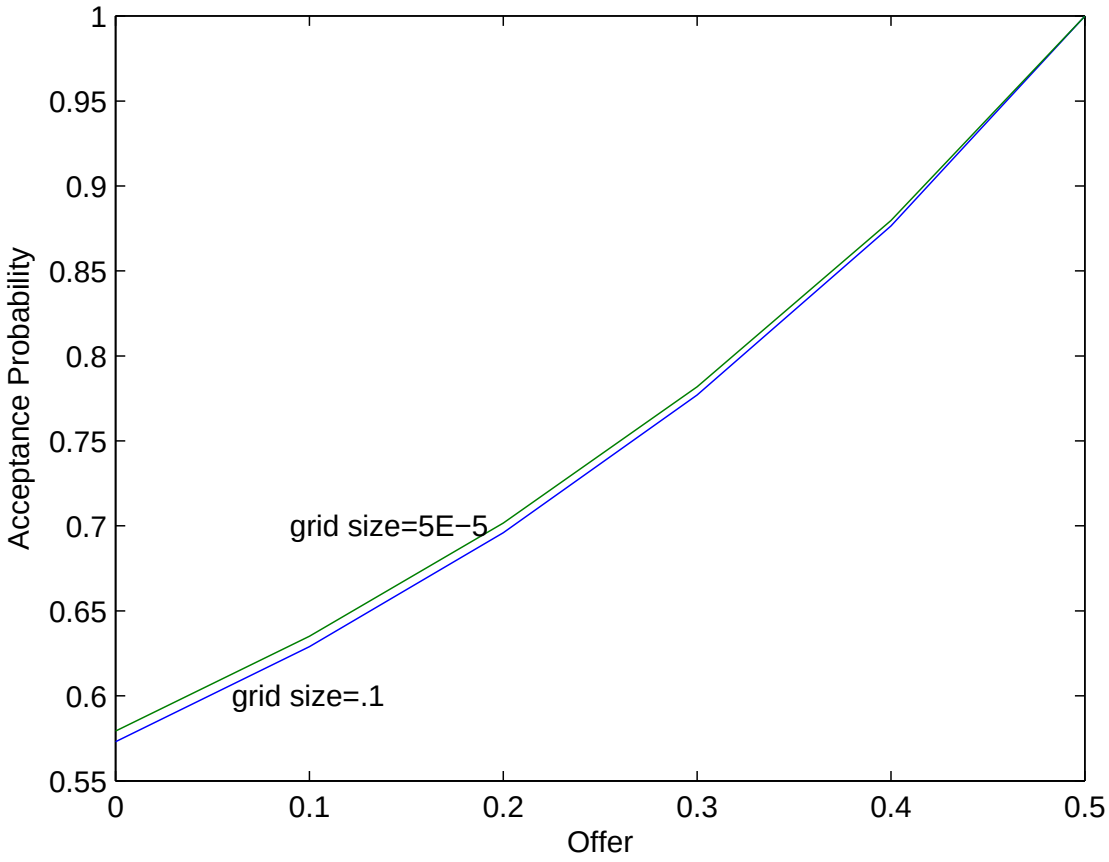

Figure 2: Probability of Acceptance: Variations in the distribution of $\bar{\lambda}$

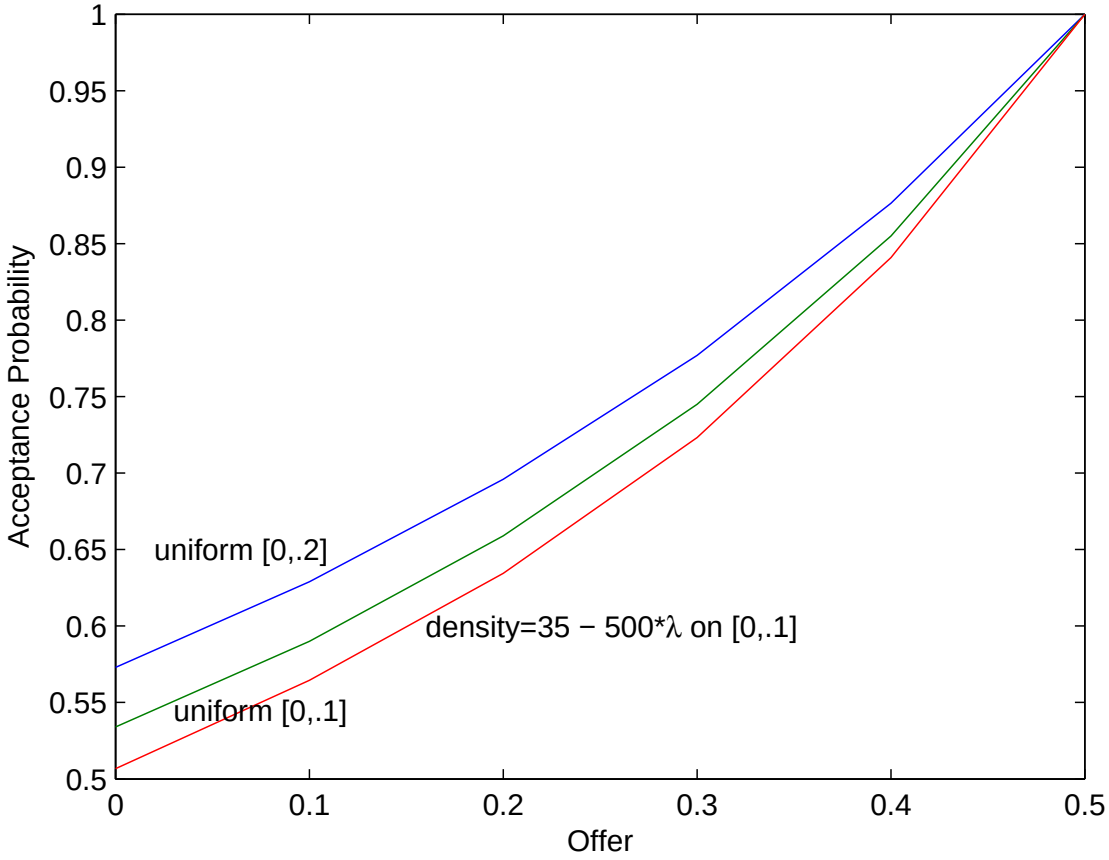

Figure 3: Probability of Acceptance: Variations in γ