

Blaming the Messenger: Notes on the Current State of Experimental Economics

Catherine Eckel* and Herbert Gintis**

March 31, 2009

1 Methodology

Ken Binmore and Avner Shaked are highly respected economists, well-known for their analytical contributions and breadth of knowledge. Moreover, they have actively participated in experimental economics for many years. However, their critique of the current state of experimental economics in general, and of Ernst Fehr and Klaus Schmidt's presentation of their theory of inequality aversion in particular, are deeply flawed. Moreover, their conception of the relationship between theory and experiment and their interpretation of the empirical evidence on other-regarding preferences are untenable.

Binmore and Shaked set the problem as follows:

Should we follow those experimental economists who seek recognition of their subject as a science by adopting the scientific standards that operate in neighboring disciplines like biology or psychology? Or should we... [treat] experimental results as just one more rhetorical tool to be quoted when convenient in seeking to convert others to whatever your own point of view may be?

Binmore and Shaked in fact identify only one aspect of scientific standards on which experimentalists are purportedly wanting, their failure to adopt "a more skeptical attitude when far-reaching claims about human behavior are extrapolated from very slender data." We maintain that experimental economics has not been faulty in this respect.

*School of Economic, Political and Policy Sciences, University of Texas at Dallas. **Santa Fe Institute and Central European University. We would like to thank Samuel Bowles, Juan Camillo Cardenas, Ernst Fehr, Barkley Rosser, and Klaus Schmidt for helpful comments, and the European Science Foundation for research support.

2 Self- and Other-regarding Behavior

Experimental economics generates data useful for assessing the strengths and weaknesses of alternative theoretical models. Neoclassical economics provides the most prominent model of individual behavior, that of the rational, self-regarding actor.¹ For this reason, much work in experimental economics has been directed towards determining the range over which the neoclassical model successfully predicts behavior. It has been determined that this range includes market-type games such as double auctions, single auctions with private values, procurement contracting, and market search (Smith, 1982; Ketcham et al., 1984; Alger, 1987; Davis and Holt, 1993; Cox et al., 1996; Cox and Oaxaca, 1996, 2000; Cason and Friedman, 2003; Sobel, 2007). A plausible generalization is that neoclassical theory depicts individual choice accurately in market-like contexts involving the interaction of many mutually anonymous agents capable of forming complete, third-party enforceable contracts, so that such other-regarding objectives as fairness and reciprocity cannot be attained or have no normative standing. As Binmore and Shaked stress, these experimental findings strongly support the continued relevance of the self-regarding actor model to many economic contexts.

However, experimental findings indicate that the neoclassical model excludes a considerable range of strategic interactions, particularly those that reproduce conditions in which normative behavior is prevalent in every-day social life. This conclusion is, in Binmore and Shaked's terminology, a "far-reaching claim," but it is one we believe amply justified by an impressive and constantly growing body of evidence.

The goal of assessing the range of applicability of neoclassical economics is rendered difficult by the fact that unless we add considerable information concerning the distribution of beliefs across players, contemporary game theory does not tell us how self-regarding agents behave (Aumann and Brandenburger, 1995), and we have no systematic model of the formation of beliefs, and hence of their distribution across game players (Gintis, 2009). However, it is plausible to assume that players will choose strategies that survive one or two rounds of backward induction.

¹We use the term "neoclassical" to mean the analysis of market exchange based on the rational, self-regarding actor model. Binmore and Shaked use the term "selfish" to mean "maximizes his own payoff," and "other-regarding" to mean "not selfish." We concur with their use of "other-regarding," but we prefer "self-regarding" to "selfish" as the use of the latter term often causes confusion because in common usage "selfish" means exclusive concern with one's own advantage, which is of course compatible with being other-regarding if an agent includes something other than his own payoff in his "advantage." Unless otherwise stated, we assume agents are "rational," by which we mean, as do Binmore and Shaked, that they maximize their preference function given their subjective priors and the constraints imposed by the situation.

With this assumption, we can assert that experimental economics has found several very important cases in which the predictions of the self-regarding actor game theoretic model are incorrect, and the failures are due precisely to the violation of the assumption of self-regarding behavior. These included the dictator game (Forsythe et al., 1994; Hoffman et al., 1994c, 1996; Eckel and Grossman, 1996; Henrich et al., 2005), the ultimatum game (Güth et al., 1982; Eckel and Grossman, 2001; Henrich et al., 2005), the prisoner's dilemma (Kiyonari et al., 2000), the trust game (Berg et al., 1995; Burks et al., 2003), the gift exchange labor market game (Fehr et al., 1997), third-party punishment game (Fehr and Fischbacher, 2004), the contract games that are the subject of the Binmore and Shaked critique, and the public goods game with punishment (Fehr and Gächter 2000, 2002).² Far from being "slender data," this is an impressive amount of data, internally consistent and pointing in the same direction .

Given the typical division of labor between experimenters and theorists in most fields of science, experimenters have no obligation, and usually no expertise, in supplying theories that are superior to the ones that they have refuted. They do so only when there are obvious alternatives that can be well-supported by straightforward variation of the experimental design. In many cases, it is not plausible to hold, as do Binmore and Shaked, that these alternatives are supported only by "slender data," and hence are worthy only of a "skeptical attitude." For instance, in many cases we have rock-solid evidence that the other-regarding behaviors observed are due to a preference for altruistic cooperation or punishment rather than "inexperience," or "inadequate incentives," to use Binmore and Shaked's favorite grounds for dismissing the evidence for the observation of other-regarding preferences.

For instance, the fact that the Responder does not choose the strictly dominant strategy in the ultimatum game has been replicated in many different social settings (Güth and Tietz, 1990; Roth et al., 1991; Camerer and Thaler, 1995; Henrich et al., 2004). Camerer (2003) provides surveys of fifteen such studies, and shows strikingly regular findings. Binmore and Shaked remark that the ultimatum game evidence does not imply other-regarding behavior because any division of the pie is a Nash equilibrium. However, the (0,0) outcome is not a Nash equilibrium, and in fact strictly positive offers are rejected with high frequency in many experiments, a behavior not compatible with the self-regarding actor model. The rejection of positive offers suggests that the Responder in this game is inequality averse, or is sacrificing to uphold a social norm, or is punishing the Proposer for being unfair. In

²Positive offers in the dictator game, rejections of positive offers in the ultimatum game, cooperation in the prisoner's dilemma, returning a positive amount in the trust game, supplying supra-minimal effort in the labor market game, positive punishment in the third party punishment game, and punishing in the final round of the public goods game with punishment, or on any round in the pure stranger treatment, all involve not playing the strictly dominant strategy.

support of the last interpretation, if the offers in an ultimatum game are generated by a computer rather than by the Proposer, and if Responders know this, low offers are less frequently rejected than in the standard ultimatum game (Blount, 1995; Bellmare et al., 2007). Similarly, if the size of the Proposer's endowment is the private information of the Proposer, rejections also fall (Croson, 1996). Moreover, in a variant of the game in which a Responder rejection leads to the Responder getting nothing but allows the Proposer to keep the share he suggested for himself, Responders rarely reject offers, and Proposers make considerably smaller (but still positive) offers (Bolton and Zwick, 1995). Given this impressive body of evidence, Binmore and Shaked's defense of the neoclassical analysis on the grounds that "all possible divisions of the money in the ultimatum game are Nash equilibrium outcomes" is transparently faulty.

The dictator game has not been studied as extensively as the ultimatum game, but a considerable range of studies based on this game also show positive contributions. As rightly noted by Binmore and Shaked, dictator game contributions are quite sensitive to differences in protocol and context. An increase in social distance leads to more selfish behavior (Hoffman et al., 1994b,d), though average contributions remain about 15% of the total pie. In addition, using a 'deserving' recipient instead of a random student subject increases giving (Eckel and Grossman, 1996). But all of these results are consistent with the general notion that other-regarding preferences draw on social norms that players infer from contextual clues encoded in game protocols.

A parallel analysis of the data concerning the final round of a public goods game with punishment permits us to conclude that altruistic punishment occurs with high frequency and the self-regarding subgame perfect equilibrium is rarely played. Binmore and Shaked's defense of the neoclassical analysis on the grounds that "full cooperation in public goods games with punishment is a Nash equilibrium outcome" is correct but irrelevant. The important point is that low contributors are frequently punished in the public goods game, even in late rounds of the game and under stranger and complete stranger treatments. This behavior is a strong indicator of other-regarding preferences.

Binmore and Shaked also argue that other-regarding behavior disappears or is greatly attenuated when the stakes are high. It is not clear why this is a criticism. We expect rational agents to adjust their behavior when the relevant parameters change. As it is more costly to realize other-regarding objectives when the stakes are high, we might expect higher stakes to reduce other-regarding behavior. Moreover, even when there is considerable attenuation, the whole range of observed behavior is theoretically important. Indeed, from a practical point of view, there are many areas of social life in which the cost of cooperating and punishing are low, yet the aggregate social benefits from such other-regarding behavior are very large

(e.g., being considerate in public settings, as well as shunning or otherwise punishing those who are not). However, in fact, behavior appears to be rather insensitive to stakes in most experiments, and often dramatically so.

For instance, there is little evidence that altruistic punishment in the ultimatum game disappears, or is even substantially attenuated, when the stakes are high (Roth et al., 1991; Hoffman et al., 1994a; Straub and Murnighan, 1995; Cameron, 1999). Slonim and Roth (1998) show that if the ultimatum game is repeated ten times with different partners each time, there is a significant, though small, tendency for rejections to decline when the stakes are very high (about ten days' wages in Slovakia), but not otherwise. Camerer and Hogarth (1999) review studies to date that test for stakes differences and find that "no replicated study" has made other-regarding behaviors "disappear purely by raising incentives."

Another study that varies stakes and experience is provided by Munier and Zaharia (2003), who conduct a within-subjects design with two different stakes levels in two countries: France and Romania. This study is an example of a new wave of experimental research that explores games such as the ultimatum game across countries, comparing play across ethnic and cultural boundaries. Like Slonim and Roth (1998) and Cameron (1999), they exploit wage differences across countries to design unusually high-stakes games. Subjects play ten ultimatum games, five with stakes of about \$7, and five with stakes raised by a factor of 50. In Romania the average wage at the time was about 10% of that in France, making the stakes quite high for the latter game. Minimum acceptable offers (MAOs) in France were around 1/3, and double the levels in Romania for both games.³ MAOs increased slightly with repetition for a given stakes-level. Note that, while the relative share demanded by the responders fell slightly across treatments, the absolute (dollar) demand increased substantially. Even in Romania, a subject who was willing to reject offers below \$1.50 when the endowment was \$7.20 in the first game would likely reject any offer below \$22 when the endowment was \$360 in the higher stakes game. Camerer (2003) notes this is a typical result arising from the high-stakes studies: relative demands fall but absolute demands rise. In addition, these results suggests there may be cultural differences in sharing norms across even developed countries, echoing the results in Henrich, et al. for more culturally distinct populations.⁴ Eckel et al. (2008) show a similar result, with substantial hetero-

³Low stakes: 34-35% in France, 16-20% in Romania; High stakes, 28-30% in France, 11-15% in Romania.

⁴ Many other cross-cultural experiments have been conducted examining differences in fairness perceptions using different games, most prominently public goods and trust games. We do not attempt a thorough survey, but examples include Fershtman and Gneezy (2001), Beard et al. (2001), Alesina and Ferrara (2002), Willinger et al. (2003), Bouckaert and Dhaene (2004), Herrmann et al. (2008) and Buchan et al. (2009). Together with the pathbreaking work of Henrich et al. (2004,2006),

generosity in average behavior across eleven villages in rural Mexico: mean amounts offered were 36-47% of an endowment of about two days' wages, and average MAOs ranged from 25% to 50% across villages.

Even the results of the dictator game have been robust to differences in stakes. The three dictator game studies in Henrich et al. (2004), p. 27, involve substantial stakes and positive offers. The percent offered to the recipient in three small scale societies yielded offers of 31, 20 and 32 percent of the endowment, and very few or none of the subjects gave zero. High school students in Houston, TX, give \$13.34 (26.7%) from a \$40 endowment in Eckel et al. (2009). Similarly, Eckel et al. (2006) report dictator game results for experiments that were conducted in rural Mexico; average giving to a stranger from another village was 26% of an endowment of two days' income, and only about 20% of the subjects gave zero.

Finally, the public goods game also has been tested at higher stakes levels. Using a scale factor of 5, Kocher et al. (2008) find no difference in contributions or punishment in a public goods game with and without punishment.

3 The Virtue of Skepticism

Binmore and Shaked's counsel of "skepticism" in the face of "slender evidence" applies to the premature acceptance of a theory, but not to offering candidates for acceptance, as Fehr and Schmidt do in proposing their model of inequality aversion. If there are anomalies in the currently accepted theory, in every branch of science with which we are acquainted, theorists creatively jump in with a variety of wide-ranging alternatives, the acceptable among which are determined by further experiment and observation. Indeed, economists need more, not fewer, speculative and creative models of human behavior. Experimentalists can use these models to determine which among the current theoretical candidates is best supported by the data, as indeed they have done in large numbers in the case of Fehr and Schmidt (1999) and competing theories, such as Engelmann and Strobel (2004).

Binmore and Shaked defend the received wisdom in ways that are either theoretically questionable or are seriously contradicted by the body of experimental evidence. Take, for instance, their argument that "there are games in which theoretical predictions based on money payoffs are not robust to small perturbations in the rationality of the players. . . experimentalists . . . cannot escape blame if they treat such fragile examples as typical." This is an unusual argument for a scientist to make. The role of experiment is not to privilege the "typical" by ignoring anomalies, but rather to determine the range over which accepted theories are valid, and

these studies illustrate the importance of cultural variation in norms of fairness in shaping individual preferences.

searching out anomalies is a standard way accomplishing this task. The anomalies revealed by experimentalists are prerequisites to developing a cogent model of human behavior because the games involved are socially ubiquitous, however “fragile.” Moreover, the Binmore and Shaked contention that we only need a small perturbation of traditional assumptions to predict the levels of cooperation found in the public goods game, firstly, is based on arguments such as Steiner (2007), in which the unique subgame perfect equilibria is attained by many rounds of recursive elimination of strongly dominated strategies, the sort of solution Binmore and Shaked rightly say will not be implemented in reality, and secondly is contradicted by the experimental public goods game evidence that there is usually a large fraction of altruistic punishers, not the “epsilon” assumed by such perturbation theories.

It is Binmore and Shaked who exhibit a lack of skepticism in defending the received wisdom in the face of contrary evidence. The self-regarding Nash equilibrium is achieved, they assert,

[in] most games with money payoffs that have a unique Nash equilibrium—provided that the payoffs are sufficiently large and the subjects have ample time for trial-and-error learning. In spite of much rhetoric to the contrary, the one-shot Prisoners’ Dilemma is a case in point.

Once again Binmore and Shaked argue in terms of the frequency of types of games (“most games”), whereas the scientific procedure is to look for anomalies, not the fraction of time the received wisdom is supported.⁵ Moreover, their assessment is arbitrary, since there is no metric in terms of which the concept of “most” games can be ascertained. Finally, Binmore and Shaked’s evidence is Ledyard (1995) and Sally (1995), surveys that do not reflect the past decade and a half of experimental evidence. For instance, the one shot prisoner’s dilemma does not support their generalization, because the study of the sequential prisoner’s dilemma illustrates that a large fraction of subjects prefer cooperating to defecting, and especially strongly so when they are guaranteed that their partner will cooperate (Kiyonari et al., 2000). In addition, Kollock (1997) allows subjects to rank possible outcomes in the game, and shows that people strongly prefer the outcome of mutual cooperation.⁶

Contradicting their own recommendations regarding proper scientific methodology, Binmore and Shaked simply assert the shoddiness of the experimental design when they dislike the findings. For instance, consider a seventeen-society study in which the understanding of the ultimatum game was a prerequisite for

⁵An excellent example of a study that systematically examines cases where theory predicts accurately and where it does not for a set of standard games is Goeree and Holt (2001).

⁶These results suggest that defection in the one-shot prisoner’s dilemma is the result of aversion to betrayal (Bohnet et al., 2008).

playing, and the results of which were replicated in a second large-scale follow-up study (Henrich et al. 2004, 2006). These ambitious research projects were carefully carried out and have received much praise from experts in the field. Nevertheless Binmore and Shaked assert, without supplying any evidence, that the results were due to “the behavior of inexperienced subjects.”⁷ Indeed, the bulk of the evidence shows little effect of experience on behavior in these games.

4 Modeling Other-regarding Preferences

Many attempts to minimize the theoretical importance of the empirical evidence for other-regarding preferences depend on the significant context dependence of such preferences, the most important being the power of payoff-irrelevant changes in the presentation of the strategic situation to experimental subjects—so-called *framing effects*. For instance Andreoni (1995) shows that, when a game is framed in terms of positive externalities, subjects exhibit a high level of cooperation, whereas when the game is framed in terms of negative externalities, a high level of defection occurs. Similarly, Charness and Rabin (2002) show that choices in dictator subgames are very different from choices in the same branch of the larger game. We propose a straightforward extension of the standard rational actor model of which context dependence is a constituent element.

We assume that an other-regarding preference, such as for fairness, reciprocity, or honesty, is an argument in the individual’s preference function that obeys the usual axioms of consumer choice in the face of changes in prices and income (Cox et al., 2008). Andreoni and Miller (2002) have shown in a dictator game in which the dictator’s gift is taxed or subsidized by the experimenter, that preferences for giving satisfy GARP, the generalized axiom of revealed preference, (Varian, 1982) which, barring corner solutions, implies that when the “price” of giving increases, the quantity of giving will fall. A similar result is shown in field experiments testing the effect of subsidies that change the price of giving to a charitable organization (Karlan and List 2007, Eckel and Grossman 2008, Huck and Rasul, 2007a,b). If this generalizes to other forms of other-regarding preferences (for the most part, the relevant experiments have not been run), then we would expect the amount of other-regarding behavior to fall as the stakes increase, provided the stakes can be interpreted as a price.⁸

⁷Note that the Henrich et al. experiments, like several subsequent studies, are not designed to test game theory, but rather to use game theory to determine which social norms are prevalent in a particular society. To criticize these researchers for recognizing that behavior is influenced by social life makes little sense. Borrowing from other social sciences, as suggested by Binmore and Shaked, is exactly what they are doing.

⁸See Andreoni et al. (2003), who “convexify” the ultimatum game, effectively changing the price

For instance, Gneezy (2005) has shown that when the price of being honest increases, the frequency of materially costly honest statements declines. On the other hand, in the standard dictator game a change in the endowment is an income rather than a price effect, so standard consumer theory with other-regarding preferences makes no prediction as to the effect of level of endowment on the amount transferred to the recipient. In the ultimatum game, an increase in the stakes raises the cost of rejecting a given fraction of the endowment, but also raises the harm inflicted on the Proposer by rejecting the offer. Thus, in this case also, consumer theory does not predict how the probability of rejection depends on the size of the endowment.

In sum, if the GARP assumption is correct, it follows that an increase in the price or shadow price of exercising an other-regarding preference will decrease the amount demanded, but a change in the stakes has an indeterminate effect on the amount demanded. We know of no exceptions to this generalization in the experimental literature. Moreover, income effects are almost always small, implying that Engel curves for other-regarding preferences are approximately linear.

The GARP assumption concerning other-regarding preferences does offer a precaution when studying the effects of game repetition. It is plausible to assume that the marginal utility of money is constant, or at least decreases more slowly than any direct argument in the utility function. It follows that if we repeatedly offer an individual a choice between consuming a good or pocketing money, the diminishing marginal utility of the good will lead the individual to increasingly value money and devalue the good in later periods of the experiment. For instance, if we expose a subject to ten repetitions of choosing between bundles consisting of chocolate cake and money, if the price of chocolate cake is held constant, the observed demand for chocolate cake is likely to decline considerably from the first to the tenth round, especially strongly if the cake must be consumed on the spot but the money may be stored for later use. Similar reasoning suggests that an other-regarding characteristic such as honesty will never be valued more, and is likely to be valued less, in later periods of a multi-round experiment. When this occurs (which is in fact rarely the case), it is simple error to interpret this as “learning the game,” when the obvious implication of traditional consumer theory is simply satiation on a consumable. One way of testing “learning the game” with self-regarding preferences versus diminishing marginal utility with other-regarding preferences is to space game repetition sufficiently that satiation effects are avoided, for instance by allowing many hours or days between each repetition.

Our second basic assumption is that the standard for valuing an other-regarding

of punishment, and find results consistent with preferences for own and other payoffs that are continuous, nonlinear, strictly convex, and nonmonotonic.

variable is governed by *social norms*, and social norms are context specific. The relative values of other-regarding variables, and their trade-offs with self-regarding variables, are context-dependent because different contexts are associated in the minds of subjects with different social values. Every social group supports a number of standard types of social interaction, each of which includes a nexus of social norms that indicate appropriate behavior for that type of interaction. We may call these standard interactions *social frames*. Social frames are identified by certain environmental cues, to which an individual attends to determine with which social context is currently appropriate. When faced with a novel social situation, a subject attends to the way the situation is framed to embrace a specific, context-dependent, preference ordering.

The norms associated with a social frame work in two ways. First, they allow subjects to form more or less accurate expectations as to the behavior of others in the social encounter. Thus, subjective priors, which are given *a priori* and unanalyzed in the standard rational choice model for an isolated individual, are variable and context dependent in a strategic interaction governed by social norms. Note that this much is true independent from the operation of other-regarding preferences. The social norms governing an interaction thus can induce a specific Nash equilibrium in an encounter with multiple equilibria (Binmore, 2005; Bicchieri, 2006).

Social norms also generate ethical standards of behavior that are highly situation specific. Thus what is fair, kind, or virtuous in one situation may be unfair, unkind, or otherwise blameworthy in another. For instance, honesty is highly valued in many social situations, but in some, such as competitive poker, the capacity to deceive is highly valued, and in polite society, disingenuous compliments are tolerated and even encouraged. Social frames in the context of other-regarding preferences thus affect the way individuals trade off among objectives. In particular, dominant social norms tend to foster a social morality in which many individuals consistently sacrifice a degree of individual material reward on behalf of common social goals (Parsons, 1937; Gintis, 2009).

The implication of these two axioms of other-regarding behavior is that "learning" and "experience" in playing a game involve not only "learning what maximizes personal material payoffs," but also "learning the norms that govern the social interaction." Similarly, "experience" with playing the game is measured by the extent to which subjects have discovered how others play the game, and hence what the appropriate norms are for playing the game. We assert that all economic experiments to date are consonant with this model, and are more plausible than alternative models that treat "learning" and "experience" in terms of the self-regarding game-theoretic prediction.

For instance, in a well known experiment, Hoffman and Spitzer (1982) found

that a Proposer's bargaining position was strengthened when both he and the Recipient were informed that the Proposer "earned" the right to the endowment rather than being "allocated" the endowment. A similar property right norm is shown by Cherry et al. (2002), in which ultimatum game offers were considerably lower when Proposers were told they had "earned" the right to the endowment, and by List and Cherry (2000), who show that Proposers who had prevailed in a contest for the right to play a high-stakes ultimatum game offer lower shares than Proposers who had lost the contest, and were obliged to play a low-stakes ultimatum game. Other experiments confirming that having the dictator or the recipient earn the endowment substantially shifts the allocation in the direction of the earner include Ruffle (1998) and Parrett (2006). Extreme allocations in the dictator game also are generated in situations where the dictator is given the option to take away money from the recipient. Bardsley (2008) and List (2007) report results of experiments where taking is an option, and show that allocations shift significantly in favor of the dictator. Krupka and Weber (2008) develop a means of eliciting social norms for a game independently of the play of the game. Their results confirm that social norms for these games shift in ways that are consistent with the observed play.

Harrison and McCabe (1996) support our model by showing that what is considered "fair" in the ultimatum game is not a fixed attribute of the individual's objective function, but rather is affected by the frequency of offers and rejections through the history of play. When the experimenters present conditions that lead to unusually low offers, responders adapt by lowering their minimum acceptable offers. We hypothesize that if subjects are presented conditions in which low offers are rare, they will in a parallel manner increase their minimum acceptable offers, although this experiment has not been carried out, to our knowledge. In each case, individuals adjust their fairness measure according to the prevailing social norm.

5 Neoclassical Economics and Other-regarding Behavior

As Binmore and Shaked note, the assumption that agents are self-regarding is rarely contradicted in laboratory experiments using market games. This fact probably accounts for the common practice among neoclassical economists, prior to the rise of game theory in the 1980's, of assuming that agents are self-regarding: nothing is gained by assuming other-regarding preferences in the context of Walrasian markets, in which each agent faces a given price system, or more generally, a given set of demand and supply curves of the other agents.

However, Binmore and Shaked argue further that other-regarding preferences present no problem for neoclassical theory.⁹ They assert:

⁹Binmore and Shaked use the term "money maximizer" several times throughout this paper with-

Behavioral economists sometimes claim that neoclassical economists hold that people are selfish. . . . But no such [selfishness] axiom appears in standard economics textbooks. . . . When utility functions of various kinds are fitted to data obtained in laboratory experiments, neoclassical economics is therefore in no danger of being refuted.

This argument is incorrect if one includes the general equilibrium model as part of neoclassical economics. The Walrasian model explicitly assumes that individual utility is a function of personal labor and personal consumption, and the fundamental theorem of welfare economics is false if this assumption is dropped. More generally, the basic neoclassical argument in favor of a market economy based on the fundamental theorem of welfare economics fails when other-regarding preferences are introduced. In this, of course other-regarding preferences are akin to other violations of Walrasian assumptions, such as spillover effects and increasing returns to scale, which complicate and temper but do not destroy the defense of market competition. A cogent defense of the market economy can be sustained assuming what we know about other-regarding preferences, but such a defense must be more nuanced, and considerably more dependent upon empirical data, than the standard defense based on the fundamental theorem of welfare economics. For instance, as Bowles (2008) and Bowles and Hwang (2008) have pointed out, on the basis of voluminous experimental and field data, there is often a trade-off between material and moral incentives, so that the optimal policy mix between moral exhortation and traditional incentives is both challenging and potentially fruitful.

Experimental economics becomes more important when we leave competitive market interactions to analyze non-market strategic interaction, for this is precisely the realm in which other-regarding behaviors are widely observed and undermine the scientific assertions and policy recommendations that obtain simply by applying neoclassical reasoning to novel, non-market, contexts. For instance, under plausible conditions, self-regarding voters will conform to the median voter model (Black, 1948; Downs, 1957), which predicts a redistribution of income towards the median voter, but generally precludes extensive redistribution from rich to poor unless the income distribution is so unequal that the median voter is considered poor. In fact, however, welfare states, with voter approval, are observed to redistribute strongly in favor of the poor in many cases (Fong et al., 2005). Thus, recognizing other-directed preferences, considerably more latitude for redistribution can be sustained. Similarly, the possibility of altruistic punishment allows effective forms of control over common pool resources that are not available when agents are self-regarding (Ostrom et al., 1992; Hayashi et al., 1999).

out definition. We infer from their usage that this term is shorthand for “self-regarding agent”, so a “non-money maximizer” is an agent with other-regarding preferences.

6 The Rationality of Backward Induction

A major claim of Binmore and Shaked is that experimental economics has not established that subjects are irrational.¹⁰ They argue this position in several ways, the most prominent of which is the fact that laboratory subjects do not implement subgame perfect equilibria does not imply that subjects are irrational. We believe the Binmore and Shaked position is correct, as argued at length in Gintis (2009). Rationality does not imply backward induction, although common knowledge of rationality (CKR) does (Aumann, 1987), and it can be shown that CKR is not a plausible epistemic condition (Gintis, 2009). As Binmore and Shaked quip, “like the zombies in horror movies that keep getting up no matter how many bullets are pumped into them, backward induction seemingly cannot be laid to rest.”

Given the widespread belief among economists that rationality implies backward induction, a credo that is perpetuated in most graduate game theory textbooks, it is useful to give a simple example showing that this is not the case.

Suppose Alice and Bob play the Prisoner’s Dilemma shown to the right, with $T > R > P > S$ and $R > 0$, repeated until one player defects or 100 rounds have been played. Backward induction implies that both players will defect in the very first round, and indeed, this is the only

	<i>C</i>	<i>D</i>
<i>C</i>	R,R	S,T
<i>D</i>	T,S	P,P

Nash equilibrium of the game. Moreover, Aumann (1987) shows that assuming common knowledge of rationality, this is the only outcome possible. However, if Alice is rational, she must have a subjective prior over the first round on which her partner will defect. Of course, if CKR holds, this prior will place probability one on her partner’s immediate defection. But if Alice merely assumes her opponent is rational, Alice’s prior is constrained only to assume her partner will defect before the final round. Alice will then choose a round to defect that maximizes her expected return given this prior. In general, this will not imply immediate defection.

More formally, suppose Alice conjectures that for $k = 1, \dots, 99$, Bob will cooperate up to round k and then defect, with probability g_k . Then, Alice will choose a round m to defect in that maximizes the expression

$$\pi_m = \sum_{i=1}^{m-1} ((i-1)R + S)g_i + ((m-1)R + P)g_m + ((m-1)R + T)(1 - G_m), \quad (1)$$

where $G_m = g_1 + \dots + g_m$. The first term in this expression represents Alice’s payoff if Bob defects first, the second term represents her payoff if Bob

¹⁰Binmore and Shaked use the term “rational” to mean “maximizes a preference function,” or equivalently, “has consistent preferences” (Savage, 1954).

and Alice defect simultaneously, and the final term is Alice's payoff if she defects first. In many cases, maximizing this expression suggests cooperating for many rounds for all plausible probability distributions. For instance, suppose g_k is uniformly distributed in the rounds $m = 1, \dots, 99$. Suppose, for concreteness, $(T, R, P, S) = (4, 2, 1, 0)$. Then, by using equation (1) it turns out that it is a best response to cooperate up to round 98. Alternatively, suppose Alice expects Bob to defect in round 1 with probability 0.95 and otherwise defect with equal probability on any round from 2 to 99. Then it is still a best response for her to defect in round 98! Clearly, the backward induction assumption is not plausible unless Alice believes her opponent is highly likely to be an obdurate backward inductor, which is equivalent to her assuming CKR.

A dilemma arises if Alice assumes Bob has the same subjective prior as herself, for then Alice may say to herself, "Bob is just as capable as I am of reasoning as above, so he will also cooperate at least up to round 98. Thus, I should set $m = 97$. But, of course, Bob also knows this, so he will surely defect in round 96, in which case I should surely defect in round 95." This reasoning leads Alice eventually to immediate defection. However this reasoning is self-contradictory. If Alice's $\{g_k\}$ distribution is reasonable, then she must use it. It is self-contradictory to use this distribution to show that it is the wrong distribution to use.

Suppose, however, that it is common knowledge that both I and my partner have the *same Bayesian priors* concerning when the other will defect. This is sometimes called *Harsanyi consistency* (Harsanyi, 1967). Then, it is obvious that we will both defect at our first opportunity because the backward induction conclusion now follows from a strictly Bayesian argument: the only prior that is compatible with mutual knowledge of common priors is defection in round 1. However, there is no plausible reason for us to assume Harsanyi consistency in this case. We conclude that *CKR is not generally a plausible assumption for rational agents*, and hence backward induction is generally not an expected behavior of rational agents.

Of course, as argued in section 2, it is reasonable to assume that rational players will engage in one or two rounds of backward induction, and hence self-regarding rational agents will implement the subgame perfect equilibrium in simple games like the one-shot ultimatum game or one round of the public goods with punishment.

7 Inequality Aversion

Binmore and Shaked hold up the case of Fehr and Schmidt's (1999) parameterization of their model of inequality aversion as an egregious example of what's wrong with experimental economics. While Fehr and Schmidt (2009) do a fine job of

defending themselves against these charges, there remain a few important points to clear up.

The Fehr-Schmidt model is faulted, among other things, for its failure to account for many influences on behavior. All models are simplifications, of course, and the Fehr-Schmidt model is no exception. The authors *intentionally* abstract from many factors of considerable importance, including the intentions of the decision makers (Falk et al., 2008), other options available to the proposer (Falk et al., 2003), or the possibility that some people value the overall payoff of the group or put particular weight on the payoff of the least well off group member (Kagel and Wolfe, 2001; Engelmann and Strobel, 2004). The remarkable thing about this very simple model is that *any* set of parameters is consistent with data from a variety of games, however Fehr and Schmidt produced them. That they are derived from a variety of data sets is commendable. The critical factor is whether these parameters hold up when confronted with other data sets. Our preference would be to estimate the parameters of the model using a variety of data sets, using the tools of experimental design and econometrics to examine the validity of the particular parameters discovered by Fehr and Schmidt as well as the overall implications of the model.

Binmore and Shaked assert that the best studies use models to predict novel phenomena and to test the relative explanatory power of competing models. Testing theory of course occupies much of the time and attention of experimentalists, and this theory has not gone untested. Four types of tests are available to assess the validity of the Fehr-Schmidt inequality aversion model and its competitors. First, experimental data can be used to estimate the parameters of the Fehr-Schmidt model, under the assumption that it is valid. We know of four such studies. Unfortunately these studies do not report a distribution of parameter values, but rather a coefficient and standard error. Ideally one would like to estimate the distribution of parameters in the population, then compare that distribution with the Fehr-Schmidt model. This has not been done.

These four studies are compared in Table 1. We first calculate the means and standard deviations of the distributions suggested by Fehr and Schmidt (note these distributions are not Gaussian or even symmetric). The first study in the table is an ultimatum game with a representative sample of Dutch adults recruited from a panel maintained for survey and experimental research (Bellmare, Kröger and van Soest 2007; see also Bellmare, Kröger and van Soest 2008). The overall sample has much larger estimated parameter values for both disadvantageous inequality (α) and advantageous inequality (β) than the Fehr-Schmidt values. However, when restricted to those with high education (university or advanced technical) and under age 35, the α parameter is not significantly different from the Fehr-Schmidt value, and the β estimate is not significantly different from zero. Goeree and Holt

(2000) report results of alternating-offer bargaining games. Their estimate of α is quite close to the Fehr-Schmidt value. They find a difference in β between proposers and responders, confirming that the environment can affect the importance of inequality as a motive for decision makers. Ellingsen and Johannessen (2004) look at the hold-up problem, and report a lower α than Fehr-Schmidt, but a value of β that is very close. Finally, while the other studies rely on choice data to estimate the parameters of the model, Daruvala uses an innovative procedure to elicit willingness-to-pay for greater equality. In her experiments, the parameters can be calculated at the individual level. She also uses university students as her subject pool, and reports average parameter values that are close to those proposed by Fehr and Schmidt. Overall, if one looks only at means, the estimates are mostly insignificantly different from Fehr-Schmidt. More notable is the fact that the estimated α parameter (or the calculated mean) is always statistically significantly different from zero, indicating support for the importance of aversion to disadvantageous inequality in particular. The estimates and significance levels of advantageous inequality are more variable.

Second, experiments can be designed to distinguish among theories. Charness and Rabin (2002) design experiments where inequality aversion and a competing theory, that of Bolton and Ockenfels (2000), are tested against self-regarding preferences, social welfare maximization and reciprocity, and propose a more general theory encompassing all of the motives. Kagel and Wolfe (2001) test the two theories in a three-person setting. Finally, Engelmann and Strobel (2004) conduct a set of three-person allocation games to test Fehr-Schmidt, Bolton-Ockenfels, and Charness-Rabin. None of the theories explains all of the behavior. Because it is more general, Charness-Rabin is better able to fit the data from a broad set of allocation games. Engelmann and Strobel (2004) provide a particularly strict test of the theories, illustrating that in settings designed to be extreme in order to explicitly to test whether behavior conforms to these theories, none explains more than a fraction of the variation in the games.

Third, as Binmore and Shaked assert, the theory can be used to predict outcomes, and the data compared with those outcomes. Fischbacher et al. (2009) do this, concluding, "A model that combines heterogeneous fairness concerns with decision errors predicts all comparative static effects of changes in competition correctly." Their Table 8 shows that the absolute predictions are quantitatively accurate in most of the games, the ultimatum game, and the game with proposer competition, but only qualitatively (relative to the ultimatum game) for the games with responder competition.

Fourth, econometric analysis of experimental data can include terms designed to capture inequality aversion, without estimating the utility function directly. For example, Ashley et al. (2009) analyze the dynamics of data from two classic pub-

lic goods games, including variables that measure deviation of own contributions from group average contributions separately if the variable is positive or negative. These terms are easily derived from a Fehr-Schmidt utility function over payoffs in terms of individual contributions to the public good. They find that contributions adjust quickly downward if they are above average, reflecting the importance of disadvantageous inequality aversion, but do not adjust if they are below average.

We conclude from these studies that accounting for the possibility of inequality aversion often improves the ability of researchers to explain data from ultimatum, public goods and related games. In some cases, the precise distribution proposed by Fehr-Schmidt gives predictions that are quantitatively accurate. More often, the standard of qualitative accuracy is attained. However, experiments designed to be strict tests of these models fail to support this or related models. Indeed, as remarked to one of us by an anonymous referee, “A half century of experimental economics has established the following fact: for every proposed theory of choice there exists an experimentalist clever enough to devise an experiment to refute it.” Social preference theories are no exception. However this fact does not invalidate the importance of the Fehr-Schmidt model for helping economists understand behavior in experimental games and in the field.

8 Conclusion

Binmore and Shaked cogently charge that experimentalists are often too quick to deem observed behavior as “irrational,” a tendency that conveniently relieves researchers from the obligation to explain human behavior. However, Binmore and Shaked appear to believe that theorists have no responsibility in explaining (as opposed to ignoring or explaining away) experimental findings. Rather, they assert that experimentalists themselves are responsible for finding an alternative to a standard theory that cannot stand up to empirical data. For instance, in response to the observation of Henrich et al. (2005) that “Experimental play often reflects patterns of interaction found in everyday life,” Binmore and Shaked reply: “If this is right, then experimentalists need to borrow ideas from social sciences other than economics to make sense of the behavior...”. We quite agree with Binmore and Shaked’s conclusion that we need to borrow ideas from other disciplines (Gintis, 2009), but the notion that it is experimentalists who are supposed to carry out this task is incorrect. “You broke it, so you go fix it” may be appropriate in some walks of life, but not for the halls of science.

REFERENCES

- Alesina, A., Ferrara, E.L., 2002. Who trusts others? *Journal of Public Economics* 85, 20–34.
- Alger, D., 1987. Laboratory tests on equilibrium predictions with disequilibrium data. *Review of Economic Studies* 54, 105–145.
- Andreoni, J., 1995. Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics* CX, 1–21.
- Andreoni, J., Castillo, M., Petrie, R., 2003. What do bargainers' preferences look like? experiments with a convex ultimatum game. *American Economic Review* 93, 672–685.
- Andreoni, J., Miller, J.H., 2002. Giving according to garp: an experimental test of the consistency of preferences for altruism. *Econometrica* 70, 737–753.
- Ashley, R., Ball, S., Eckel, C., 2009. Motives for giving: a reanalysis of two classic public goods experiments. Working paper, Center for Behavioral and Experimental Economic Science.
- Aumann, R.J., 1987. Correlated equilibrium and an expression of Bayesian rationality. *Econometrica* 55, 1–18.
- Aumann, R.J., Brandenburger, A., 1995. Epistemic conditions for Nash equilibrium. *Econometrica* 65, 1161–1180.
- Bardsley, N., 2008. Dictator game: altruism or artefact? *Experimental Economics* 11, 122–133.
- Beard, T.R., Beil, R.O., Mataga, J., 2001. Reliant behavior in the united states and japan. *Economic Inquiry* 39, 270–279.
- Bellmare, C., Kröger, S., van Soest, A., 2007. Preferences, intentions, and expectations: a large-scale experiment with a representative subject pool. IZA (Institute for Labor Economics) Working Paper No. 3022.
- Bellmare, C., Kröger, S., van Soest, A., 2008. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica* 76, 815–839.
- Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10, 122–142.
- Bicchieri, C., 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press, Cambridge.
- Binmore, K.G., 2005. *Natural Justice*. Oxford University Press, Oxford.
- Black, D., 1948. On the rationale of group decision-making. *Journal of Political Economy* 56, 23–34.

- Blount, S., 1995. When social outcomes aren't fair: the effect of causal attributions on preferences. *Organizational Behavior & Human Decision Processes* 63, 131–144.
- Bohnet, I., Greig, F., Herrmann, B., Zeckhauser, R., 2008. Betrayal aversion. *American Economic Review* 98, 294–310.
- Bolton, G.E., Ockenfels, A., 2000. A theory of equity, reciprocity and competition. *American Economic Review* 90, 166–193.
- Bolton, G.E., Zwick, R., 1995. Anonymity versus punishment in ultimatum games. *Games and Economic Behavior* 10, 95–121.
- Bouckaert, J., Dhaene, G., 2004. Inter-ethnic trust and reciprocity: results of an experiment with small business entrepreneurs. *European Journal of Political Economy*, 20(4), 869–886 20, 869–886.
- Bowles, S., 2008. Policies designed for self-interested citizens may undermine “the moral sentiments”: evidence from economic experiments. *Science* 320, 1605–1609.
- Bowles, S., Hwang, S-H., 2008. Social preferences and public economics: mechanism design when preferences depend on incentives. *Journal of Public Economics* 92, 1811–1820.
- Buchan, N., Grimalda, G., Wilson, R., Brewer, M., Fatas, E., Foddy, M., 2009. *Globalization and human cooperation*. Rice University.
- Burks, S.V., Carpenter, J.P., Verhoogen, E., 2003. Playing both roles in the trust game. *Journal of Economic Behavior and Organization* 51, 195–216.
- Camerer, C., 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton, NJ.
- Camerer, C., Hogarth, R., 1999. The effects of financial incentives in experiments: a review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19, 7–42.
- Camerer, C., Thaler, R., 1995. Ultimatums, dictators, and manners. *Journal of Economic Perspectives* 9, 209–219.
- Cameron, L.A., 1999. Raising the stakes in the ultimatum game: experimental evidence from indonesia. *Economic Inquiry* 37, 47–59.
- Cason, T., Friedman, D., 2003. Buyer search and price dispersion: a laboratory study. *Journal of Economic Theory* 112, 232–260.
- Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117, 817–869.
- Cherry, T., Frykblom, P., Shogren, J., 2002. Hardnose the dictator. *American Economic Review* 92, 1218–1221.

- Cox, J.C., Friedman, D., Sadiraj, V., 2008. Revealed altruism. *Econometrica* 76, 31–69.
- Cox, J.C., Isaac, R.M., Cech, P., Cann, D., 1996. Moral hazard and adverse selection in procurement contracting. *Games and Economic Behavior* 17, 147–176.
- Cox, J.C., Oaxaca, R.L., 1996. Is bidding behavior consistent with bidding theory for private value auctions. In: Isaac, R.M. (Ed.), *Research in Experimental Economics VI*. Jai Press, Greenwich, CN.
- Cox, J.C., Oaxaca, R.L., 2000. Good news and bad news: search from unknown wage offer distributions. *Experimental Economics* 2, 197–225.
- Croson, R., 1996. Information in ultimatum games: an experimental study. *Journal of Economic Behavior & Organization* 30, 197–212.
- Daruvala, D., 2010. Would the right social preference model please stand up? *Journal of Economic Behavior and Organization* Forthcoming.
- Davis, D.D., Holt, C.A., 1993. *Experimental Economics*. Princeton University Press, Princeton, NJ.
- Downs, A., 1957. *An Economic Theory of Democracy*. Harper & Row, Boston.
- Eckel, C., Grossman, P., 1996. Altruism in anonymous dictator games. *Games and Economic Behavior* 16, 181–191.
- Eckel, C., Grossman, P., 2001. Chivalry and solidarity in ultimatum games. *Economic Inquiry* 39, 171–188.
- Eckel, C., Grossman, P., 2008. Subsidizing charitable contributions: a natural field experiment comparing matching and rebate subsidies. *Experimental Economics* 11, 234–252.
- Eckel, C., Grossman, P., Johnson, A., de Oliveira, A., Rojas, C., Wilson, R., 2009. Social norms of sharing in high school: teen giving in the dictator game. Working paper, Center for Behavioral and Experimental Economic Science, University of Texas at Dallas.
- Eckel, C., Johnson, A., Thomas, D., 2006. Measurement of preferences in socio-economic surveys: early results from the mexican family life survey. Presentation, Allied Social Science Association.
- Eckel, C., Johnson, A., Candelo Londono, N., 2008. The ultimatum game in 11 mexican villages: altruism, risk preferences and beliefs. Working paper, Center for Behavioral and Experimental Economic Science, University of Texas at Dallas.
- Ellingsen, T., Johannessen, M., 2004. Promises, threats and fairness. *The Economic Journal* 114, 397–420.

- Engelmann, D., Strobel, M., 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review* 94, 857–869.
- Falk, A., Fehr, E., Fischbacher, U., 2003. On the nature of fair behavior. *Economic Inquiry* 41, 20–26.
- Falk, A., Fehr, E., Fischbacher, U., 2008. Testing theories of fairness—intentions matter. *Games and Economic Behavior*, 62 1, 287–303 62, 287–303.
- Fehr, E., Fischbacher, U., 2004. Third party punishment and social norms. *Evolution & Human Behavior* 25, 63–87.
- Fehr, E., Gächter, S., 2000. Cooperation and punishment. *American Economic Review* 90, 980–994.
- Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. *Nature* 415, 137–140.
- Fehr, E., Gächter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: experimental evidence. *Econometrica* 65, 833–860.
- Fehr, E., Schmidt, K., 2009. What’s wrong with binmore and shaked’s criticism of inequity aversion? *Journal of Economic Behavior and Organization* ??, ??–??
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114, 817–868.
- Fershtman, C., Gneezy, U., 2001. Discrimination in a segmented society: an experimental approach. *Quarterly Journal of Economics* 116, 351–377.
- Fischbacher, U., Fong, C.M., Fehr, E., 2009. Fairness, errors and the power of competition. *Journal of Economic Behavior and Organization* in press.
- Fong, C.M., Bowles, S., Gintis, H., 2005. Reciprocity and the welfare state. In: Gintis et al. (2005).
- Forsythe, R., Horowitz, J., Savin, N.E., Sefton, M., 1994. Replicability, fairness and pay in experiments with simple bargaining games. *Games and Economic Behavior* 6, 347–369.
- Gintis, H., 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton University Press, Princeton, NJ.
- Gintis, H., Bowles, S., Boyd, R., Fehr, E., 2005. *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life*. MIT Press, Cambridge.
- Gneezy, U., 2005. Deception: the role of consequences. *American Economic Review* 95, 384–394.
- Goeree, J., Holt, C.A., 2000. Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games. *European Economic Review* 44 44, 1079–1089.

- Goeree, J., Holt, C.A., 2001. Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review* 91, 1402–1422.
- Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3, 367–388.
- Güth, W., Tietz, R., 1990. Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology* 11, 417–449.
- Harrison, G., McCabe, K., 1996. Expectations and fairness in a simple bargaining experiment. *International Journal of Game Theory* 25, 303–327.
- Harsanyi, J.C., 1967. Games with incomplete information played by Bayesian players, Parts i, ii, and iii. *Behavioral Science* 14, 159–182, 320–334, 486–502.
- Hayashi, N., Ostrom, E., Walker, J., Yamagishi, T., 1999. Reciprocity, trust, and the sense of control: a cross-societal study. *Rationality and Society* 11, 27–46.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-scale Societies*. Oxford University Press, Oxford.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., 2005. Economic man’ in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences* 28, 795–815.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., and Ziker, J., 2006. Costly punishment across human societies. *Science* 312, 1767–1770.
- Herrmann, B., Thöni, C., Gächter, S., 2008. Anti-social punishment across societies. *Science* 319, 1362–1367.
- Hoffman, E., McCabe, K., Smith, V.L., 1994a. On expectations and the monetary stakes in ultimatum games. *International Journal of Game Theory* 7, 289–302.
- Hoffman, E., McCabe, K., Shachat, K., Smith, V.L., 1994b. Preference, property rights and anonymity in bargaining games. *Games and Economic Behavior* 7, 346–380.
- Hoffman, E., McCabe, K., Shachat, K., Smith, V.L., 1994c. Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior* 7, 346–380.
- Hoffman, E., McCabe, K., Smith, V.L., 1994d. Social distance and other-regarding behavior in dictator games. *American Economic Review* 86, 653–660.
- Hoffman, E., McCabe, K., Smith, V.L., 1996. Social distance and other-regarding behavior in dictator games. *American Economic Review* 86, 653–660.

- Hoffman, E., Spitzer, M.L., 1982. The coase theorem: some experimental tests. *Journal of Law and Economics* 25, 73–89.
- Huck, S., Rasul, I., 2007a. Comparing charitable fundraising schemes: evidence from a natural field experiment. Else Working Paper #274, Department of Economics, University College London.
- Huck, S., Rasul, I., 2007b. Testing consumer theory in the field: private consumption versus charitable goods. Else Working Paper #275, Department of Economics, University College London.
- 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.
- Karlan, D., List, J., 2007. Does price matter in charitable giving? evidence from a large-scale natural field experiment. *American Economic Review* 97, 1774–1793.
- Ketcham, J., Smith, V.J., Williams, A.W., 1984. A comparison of posted-offer and double auction pricing institutions. *Review of Economic Studies* 51, 595–614.
- Kiyonari, T., Tanida, S., Yamagishi, T., 2000. Social exchange and reciprocity: confusion or a heuristic? *Evolution and Human Behavior* 21, 411–427.
- Kocher, M.G., Martinsson, P., Visser, M., 2008. Does stake size matter for cooperation and punishment? *Economics Letters* 99, 508–511.
- Kollock, P., 1997. Transforming social dilemmas: group identity and cooperation. In: Danielson, P. (Ed.), *Modeling Rational and Moral Agents*. Oxford University Press, Oxford.
- Krupka, E., Weber, R., 2008. Identifying social norms using coordination games: why does dictator game sharing vary? IZA discussion paper No. 3860. Institute for Labor Economics, University of Bonn.
- Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.), *The Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ, pp. 111–194.
- List, J., 2007. On the interpretation of giving in dictator games. *Journal of Political Economy* 115, 482–493.
- List, J., Cherry, T., 2000. Learning to accept in ultimatum games: evidence from an experimental design that generates low offers. *Experimental Economics* 3, 11–29.
- Munier, B., Zaharia, C., 2003. High stakes and acceptance behavior in ultimatum bargaining: a contribution from an international experiment. *Theory and Decision* 53, 187–207.

- Ostrom, E., Walker, J., Gardner, R., 1992. Covenants with and without a sword: self-governance is possible. *American Political Science Review* 86, 404–417.
- Parrett, M., 2006. An analysis of the determinants of tipping behavior: a laboratory experiment and evidence from restaurant tipping. *Southern Economic Journal* 73, 489–514.
- Parsons, T., 1937. *The Structure of Social Action*. McGraw-Hill, New York.
- Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in jerusalem, ljubljana, pittsburgh, and tokyo: an experimental study. *American Economic Review* 81, 1068–1095.
- Ruffle, B.J., 1998. More is better, but fair is fair: tipping in dictator and ultimatum games. *Games and Economic Behavior* 23, 247–265.
- Sally, D., 1995. Conversation and cooperation in social dilemmas. *Rationality and Society* 7, 58–92.
- Savage, L.J., 1954. *The Foundations of Statistics*. John Wiley & Sons, New York.
- Slonim, R., Roth, A.E., 1998. Learning in high stakes ultimatum games: an experiment in the slovak republic. *Econometrica* 66, 569–596.
- Smith, V., 1982. Microeconomic systems as an experimental science. *American Economic Review* 72, 923–955.
- Sobel, J., 2007. Do markets make people selfish? University of California at San Diego.
- Steiner, J., 2007. A trace of anger is enough: on the enforcement of social norms. *Economic Bulletin* 8, 1–4.
- Straub, P.G., Murnighan, J.K., 1995. An experimental investigation of ultimatum game: common knowledge, fairness, expectations, and lowest acceptable offers. *Journal of Economic Behavior and Organization* 27, 345–364.
- Varian, H.R., 1982. The nonparametric approach to demand analysis. *Econometrica* 50, 945–972.
- Willinger, M., Keser, C., Lohmann, C., Usunier, J-C., 2003. A comparison of trust and reciprocity between france and germany: experimental investigation based on the investment game. *Journal of Economic Psychology* 24, 447–466.

c:\Papers\Reply to Binmore and Shaked\Template.tex March 31, 2009

<i>Study and Population</i>	Mean Parameter Values	
	α	β
Fehr and Schmidt (1999)		
Calibration (calculated from reported distributions)	0.85	0.315
Standard Deviation	(1.119)	(0.252)
Bellemare, Kröger, and van Soest ^a		
Population sample, Dutch adults	1.892	0.801
	(0.640)	(0.921)
High education and below 35	1.018	-0.271
	(0.325)	(0.503)
Goeree and Holt (2000) ^b		
Student Proposers	0.84	0.66
	(0.160)	(0.080)
Student Responders	0.84	0.12
	(0.160)	(0.020)
Ellingsen and Johannessen (2004) ^c		
Student Subjects	0.31	.36
95% Confidence Interval	(0-1.2)	(0.25)-0.5
Daruvala (2009) ^d		
Student subjects	0.952	0.344
Standard deviation	(1.039)	(0.294)

Table 1: Parameter estimates of the Fehr and Schmidt (1999) model of inequality aversion. Standard errors in parentheses, unless otherwise indicated. (a) Specification included variables to measure intentions, and is embedded in a logit decision framework; (b) Specification included logit decision framework; a common parameter was estimated; (c) Estimated from graphs; (d) Mean and standard deviation value calculated by the author from individual data.