Journal of Economic Behavior & Organization 73 (2010) 101–108

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Discussion

On inequity aversion: A reply to Binmore and Shaked[☆]

Ernst Fehr[a,*], Klaus M. Schmidt[b]

[a] Institute for Empirical Research in Economics, University of Zurich, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland
[b] Department of Economics, University of Munich, Ludwigstrasse 28, D-80539 Munich, Germany

ARTICLE INFO

Article history:
Received 8 December 2008
Accepted 6 February 2009

JEL classification:
B41
C90

Keywords:
Experiments
Other-regarding preferences
Inequity aversion

ABSTRACT

In this paper we reply to Binmore and Shaked's criticism of the Fehr–Schmidt model of inequity aversion. We put the theory and their arguments into perspective and show that their criticism is not substantiated. Finally, we briefly comment on the main challenges for future research on social preferences.

1. Introduction

The theory of inequity aversion (Fehr and Schmidt, 1999) was developed to solve a puzzle: why do most experimental subjects behave very selfishly in some games, e.g. exploiting their bargaining power in competitive market games and free-riding in the final periods of voluntary cooperation games, while demonstrating rather "fair" behavior in other games, e.g. in bilateral bargaining games, in trust games, and in public good games with punishments? In Fehr and Schmidt (1999, FS 1999 in the following), we searched for a common principle that can explain this contradictory evidence. We wanted a simple theory, one that can be used as a tractable tool in more complicated models and that yields quantitative, testable predictions.

The theory of inequity aversion proposed in FS (1999) has triggered a lively debate. The theory has been applied to many different experimental games, it has been generalized and put on an axiomatic foundation, and it has been tested against other notions of fairness or reciprocity. This debate was often critical, but always fair and to the point.

In a current paper Binmore and Shaked (2009) criticize FS (1999) and some of our subsequent work on inequity aversion.[1] They purport to use our work as a "case study" to illustrate more general methodological issues. The fact is, however, that they are hitting a straw man. In this paper we will show that their criticism is unfounded.[2]

We are going to show this by proceeding as follows. In Section 2 we sketch the background of our 1999 paper to put it into perspective. In Section 3, we discuss in which sense we "calibrated" our model in FS (1999). In Section 4 we reconsider the four games to which we applied the theory of inequity aversion in FS (1999) and show that Binmore and Shaked's critique of

☆ Klaus Schmidt gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft through SFB/TR 15. We also thank Bjorn Bartling, Urs Fischbacher, Georg Gebhardt, Larry Samuelson and an anonymous referee for useful comments.

* Corresponding author.
E-mail addresses: efehr@iew.uzh.ch (E. Fehr), klaus.schmidt@LMU.de (K.M. Schmidt).

[1] As some readers may know, Shaked's criticism of our work dates back to what he called a "pamphlet" entitled "The Rhetoric of Inequity Aversion"

0167-2681/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2009.12.001

our analysis is unfounded. In Section 5, we discuss the difference between the distribution of preferences used in FS (1999) and that in Fehr and Schmidt (2004), Fehr et al. (2007) and Fehr et al. (2008) (the three "contract papers"), and show that Binmore and Shaked's criticism is not substantiated. Finally, in Section 6 we deal with Binmore and Shaked's interpretation of the data in the contract papers.

We conclude in Section 7 with a few general remarks on the future tasks in the domain of social preference research. We anticipate that a complete characterization of the distribution of different social preference types in the population may introduce so much complexity at the individual level that models that attempt to capture this complexity may become analytically intractable. For this reason, a simple model such as the theory of inequity aversion may still be useful, even though there is evidence that it does not provide a full description of other-regarding preferences.

2. Putting the theory of inequity aversion into perspective

A substantial amount of experimental evidence has been accumulated over the last 30 years that indicates that many subjects do not behave in a purely selfish, money-maximizing way (Camerer, 2003). Even Binmore and Shaked (2009, p. 120–121), who remained sceptical about the existence of social preferences for a long time, regard this evidence now "as an informal proof that such preferences exist". When we wrote our QJE paper in 1997–1998 the facts already clearly indicated that other-regarding preferences play an important role in many experimental games. However, at the time it was common practice in many experimental papers to offer ex post, situation dependent, explanations of the observed behavior that referred to altruism, spite, status seeking, or some particular social norm, but each explanation told us very little beyond the particular experiment for which it was invented. Furthermore, the existence of other-regarding preferences was difficult to reconcile with the experimental evidence on some other experimental games. For example, in voluntary contribution public good games and in many competitive market games the majority of subjects converged towards outcomes that were predicted by the standard self-interest approach. The question thus was whether there is a general model that gives a unifying explanation of both the evidence that suggested the existence of other-regarding preferences and the evidence that seemed to suggest the absence of other-regarding preferences. FS (1999) was one of the first attempts to develop such a model.[3]

In FS (1999) we assume that people are heterogeneous: some people suffer from linear inequity aversion (to different degrees) while others are purely self-interested. The bulk of our paper is concerned with stating, proving and interpreting five theoretical propositions that provide qualitative insights into the outcomes generated by the interaction between self-interested and inequity averse players in several prominent experimental games.

In particular, the propositions offer a solution to the puzzle that despite the known existence of a large share of other-regarding players certain market games generate very unfair distributions of the gains from trade that closely resemble the predictions of the standard self-interest model that assumes that *all* players are completely selfish. Thus, our paper helps us to understand why the self-interest model predicts so well in these market games even though the assumptions of the model are wrong.[4] We also show that a small minority of selfish players in a one-shot public good game may suffice to generate a

(Shaked, 2005). Shaked distributed this pamphlet via email to more than 300 members of the game theory community and posted it on his website and at SSRN—explicitly without intending to publish it in a refereed journal. Fehr and Schmidt (2005) wrote a reply rapidly, followed by a revised version of the pamphlet (Shaked, 2005). Shaked's pamphlet has raised some confusion in the profession about the contribution of the theory of inequity aversion. Putting the offensive and polemical rhetoric aside, Shaked (2005) made three main accusations against FS (1999) and against our invited lecture at the World Congress of the Econometric Society (Fehr and Schmidt, 2003):

1. Shaked claimed that the conclusions drawn from our Proposition 4 are *"false"*, and that this *"seemingly minor mistake is crucial for the analysis of the data"* (*p.* 11).
2. He claimed that our analysis of the market game with proposer competition and the market game with responder competition does not show that the equilibrium in these two games is close to the competitive outcome even if the population is highly inequity averse. He argued *"that the first game is logically unsuitable to demonstrate (this) point and that the second game does not show it"* (p. 15).
3. He claimed that we fail to calibrate our model by using data from Ultimatum Games and that we do not explain the experimental observations in four other games using this calibration.

Binmore and Shaked have since withdrawn the first two points. In our reply (Fehr and Schmidt, 2005) we showed that the first accusation is false and results from an error in Shaked's analysis. Shaked may have been confused by a typo (">" instead of "$\geq$") in the statement (but not the proof) of Proposition 4. Shaked dropped this charge in the revised version of the pamphlet (Shaked, 2005).
The second point has now been withdrawn by Binmore and Shaked (2009). They conclude in their discussion of the two *"auctioning games"* (the market game with proposer and with responder competition that we considered in FS 1999): *"Fehr and Schmidt's inequity aversion model is no worse as a predictor of the two auctioning games than the money-maximizing model."* The money-maximizing model predicts the competitive equilibrium outcome, and so does the inequity aversion model. This point can thus be put aside as well.
However, Binmore and Shaked did not give up on the last point but extended their original criticism to three other papers (the *"contract papers"*) that we had written in the meantime. We will show in this comment that the third point and the new criticisms are also ill-founded. We hope that this settles the remaining point of the controversy, enabling us – and the profession – to devote our time to more productive enterprises.

[2] We do not comment on the more general methodological points raised in the first sections of Binmore and Shaked (2009). For a more general discussion of these issues, we refer the reader to Eckel and Gintis (in press). See also Schmidt (2009) for a methodological discussion of the role of experiments for economic theory without reference to the current controversy.

[3] Other attempts include, Levine (1998), Bolton and Ockenfels (2000), and Falk and Fischbacher (2006). See Fehr and Schmidt (2006) for a survey.

[4] See also Dufwenberg et al. (2008) who established this result more generally in a general equilibrium framework.

unique equilibrium in which all players free ride completely. This result can help us understand why we frequently observe very little cooperation in the final periods of public good games.

Our theoretical approach differs fundamentally from the alternative approach that is proposed by Binmore and Shaked (2009, p. 120–121) who favor an explanation of "the behavior of subjects in terms of social norms." They "think it likely that people enter the laboratories with a variety of social norms, one of which is triggered by the manner in which the experiment is framed. If the resulting behavior is close to a Nash equilibrium of the game (as in the Ultimatum Game), then the social norm is stabilized in the laboratory environment. If it is not (as in the Prisoners' Dilemma), then the subjects' behavior moves towards a Nash equilibrium."

One important problem with this alternative approach is that the set of social norms that affect people's behavior is not defined. A second important problem is that nowhere do Binmore and Shaked provide a stringent theory that explains how a particular framing of a situation selects the behaviorally relevant social norm. For this reason they have complete freedom to rationalize any result ex post that is compatible with a Nash equilibrium by choosing the appropriate social norm that explains the behavior.

In contrast, our theory stipulates concrete preferences which often enable us to derive point predictions and clear comparative static results. The ability to provide point predictions is important because it enhances our knowledge about social preferences by facilitating the collection of data that refute the theory. Such deviations from the theory are informative and tell us something about other forms of social preferences or other behavioral forces. And they give rise to new and better theories.

Let us illustrate the difference between Binmore and Shaked's approach and ours with an example that involves the bilateral ultimatum game and the market game with proposer competition.[5] Roth et al. (1991) have shown that across four different countries relatively egalitarian outcomes occur in the ultimatum game while in a market game with proposer competition the responder appropriates literally the whole bargaining pie. In the bilateral ultimatum game any distribution of the available money between the proposer and the responder can be supported as a Nash equilibrium. Likewise, in the market game with two or more competing proposers any division of the money between the responder and one of the proposers can be supported as a Nash equilibrium.[6] Thus, in the absence of a theory that constrains the set of social norms, *any* distribution can be rationalized by Binmore and Shaked's alternative approach. Moreover, Binmore and Shaked's approach is also capable of rationalizing *any* comparative static differences between the ultimatum game and market games with proposer competition by arbitrarily invoking different social norms across the two games. However, a theoretical approach that is consistent with any outcome explains nothing.

Contrast this with our theory of inequity aversion which makes much stronger predictions: there is a unique subgame perfect equilibrium outcome in the market game in which the responder receives the whole pie.[7] This outcome prevails in equilibrium regardless of the distribution of inequity aversion among the players. Thus, the theory makes a point prediction, and it predicts a stark comparative static result. It shows that a proposer will never offer more than 50% of the pie in the ultimatum game, while the equilibrium offer of at least two proposers is equal to one in the market game with proposer competition. If this prediction were refuted by the experimental evidence, it would demonstrate that other behavioral forces not captured by the FS model of inequity aversion play a role in this situation. Our approach thus favors cumulative learning from experiments and contrasts sharply with an approach that is free to invoke a plethora of social norms to rationalize different outcomes.[8]

3. Calibrating the model of inequity aversion

Binmore and Shaked do not dispute our theoretical results. They are concerned with Section V of FS (1999), in which we establish that the *same* distribution of types consistent with the experimental evidence in the ultimatum game is also largely consistent with the experimental evidence in the other four games that we consider. In FS (1999) we are very explicit about the purpose of this exercise:

> "In this section we examine whether the distribution of parameters that is consistent with experimental observations in the ultimatum game is consistent with the experimental evidence from the other games. ... The objective is ... to

[5] In the market game with proposer competition $n-1$ proposers ($n > 2$) simultaneously propose a distribution of the pie of size 1 to a single responder who decides whether to accept or reject the highest offer. If there are multiple highest offers, one of them is selected at random and the responder decides whether to accept this offer.

[6] The following strategy combination constitutes a Nash equilibrium for selfish subjects in the market game with proposer competition. The responder accepts the highest offer if and only if it is below or equal to a threshold y with $0 < y < 1$. Suppose that one proposer offers exactly y. Given this offer and the strategy of the responder it is optimal for all other proposers to offer y as well. Furthermore, given that all proposers offer y, it is optimal for the responder to follow his strategy, that is, to accept the highest offer if and only if it is less than or equal to y. Hence, any y in [0,1] can be supported as a Nash equilibrium outcome, even though, for any y < l, the equilibrium is not subgame perfect.

[7] See FS (1999, Proposition 2).

[8] One of the important and sometimes overlooked aspects of traditional self-interest theory is that it often enables researchers to make sharp predictions. In fact, many insights into the nature and the existence of social preferences would not have been possible if the standard theory would not make precise point predictions that were clearly refuted by the data. The self-interest approach does, of course, not always make sharp predictions (e.g. in repeated games). However, it makes sharp predictions sufficiently often to render it useful. The same argument applies to our model.

offer a first test for whether there is a chance that our theory is consistent with the *quantitative* evidence from different games. Admittedly, this test is rather crude." (p. 843)

"Clearly, the above computations provide only rough evidence in favor of our model. To rigorously test the model, additional experiments have to be run. We would like to suggest a few variants of the experiments discussed so far that would be particularly interesting: ..." (p. 846)

Binmore and Shaked suggest, however, that the purpose of this exercise is much more ambitious and that we claim to have identified a *unique* distribution of preferences consistent with the data in the ultimatum game. However, as the following quote shows, this is not the case:

"Table III suggests a simple discrete distribution of α_i and β_i. We have chosen this distribution because it is consistent with the large experimental evidence we have on the ultimatum game ..." (FS, 1999, p. 843, emphasis added).

Thus, we state very clearly that there are many degrees of freedom in choosing a distribution consistent with the ultimatum game. At no point do we claim that the ultimatum game data uniquely identify this distribution. If Binmore and Shaked prefer to call our exercise a "parameterization" then this is fine with us.[9]

4. Explaining the evidence of the four other games

Binmore and Shaked discuss whether FS (1999) is able to explain the evidence in "the four other games". As already mentioned in Footnote 1, Binmore and Shaked now acknowledge that the "inequity aversion model is no worse as a predictor of the two auctioning games than the money-maximizing model". Both models predict the competitive outcome observed in these experiments. Concerning the public good game without punishment, they claim that we "potentially misrepresent the extent to which a money-maximizing model can explain the data in Public Good Games without Punishment". There must be a misunderstanding. We say explicitly:

"The striking fact revealed by Table II is that ... the vast majority of subjects play the equilibrium strategy of complete free riding. ... In view of the facts presented in Table II, it seems fair to say that the standard model 'approximates' the choices of a big majority of subjects rather well."

Binmore and Shaked also point out that the experimental evidence is not clear cut with regard to the public goods game without punishment. There are many different papers that use different treatments and report varied results. Depending on the papers selected, one gets larger or smaller deviations from the prediction that all players free-ride completely. We agree. This is why we included Table II (FS 1999, p. 838) with a long list of public good games and the different results obtained. We only compare our theoretical results to the average experimental result across all of these papers. This may have been too simple for a rigorous test of our model, but we attempted only a first "admittedly rather crude" test of "whether there is a chance that our theory is consistent with the quantitative evidence from different games". In any case, Binmore and Shaked acknowledge that the model of inequity aversion "is no worse than the money-maximizing model in the case of the Public Good Game without Punishment". However, while the money-maximizing model relies on the counterfactual assumption that all players only care about money, our model helps us understand why free-riding is so prevalent in the final period of repeated public goods experiments, even though many subjects have other-regarding preferences (Camerer, 2003) and are thus – in principle – willing to cooperate (Andreoni, 1995; Fischbacher et al., 2001; Kocher et al., 2008).

Finally, they discuss the public good game with punishment and point out that the cooperation rates obtained differ depending on whether a stranger or a partner design is used. They criticize that we use the model of inequity aversion to predict the experimental results of the partner treatment that showed higher cooperation rates than the stranger treatment. This does not really matter from the point of view of the theory, however. As we show in Proposition 5, if there is a sufficiently large group of "conditionally cooperative enforcers", there is a continuum of equilibria with contribution levels ranging from 0 to the entire initial endowment. Thus, even if subjects contributed significantly less in the stranger treatment, this is not inconsistent with the model.

5. Keeping the distribution constant

Binmore and Shaked accuse us of not using the same distribution that we used in FS (1999) in the three contract papers, Fehr and Schmidt (2004), Fehr et al. (2007) and Fehr et al. (2008), in the following referred to as FS (2004), FKS (2007) and FKS (in press). "Fehr and Schmidt ... feel free to change the parameters of their model when new data needs to be accommodated." (Binmore and Shaked, 2009, p. 120–121). This is not true because we used the same distribution of parameters across all

[9] A lot of this debate is about semantics. Binmore and Shaked allege several times that we claim to have "estimated" the distribution of parameters. We never claim anywhere in FS (1999) that we estimated preferences parameters, nor do we utilize the term "calibration" in this paper. This term is first used in FS (2003), where we summarized what we did in our earlier paper in two sentences. It seems to us that the term "calibration" is rather cautious. We are not aware of any calibration exercise in economics where the term "calibration" is used in the sense of finding a *unique* set of parameters that is consistent with the data.

Table 1
Distribution of preference parameters in FS (1999) and the three contract games.

FS (1999)		FS (2004), FKS (2007), FKS (2008)	
Types	Share in population	Types	Share in population
$\alpha=0$, $\beta=0$	30%	$\alpha=0$, $\beta=0$	60%
$\alpha=0.5$, $\beta=0.25$	30%		
$\alpha=1$, $\beta=0.6$	30%	$\alpha=2$, $\beta=0.6$	40%
$\alpha=4$, $\beta=0.6$	10%		

three contract games although these games differ significantly from each other. If we had intended to just fit the parameters to the data of each game it would have been far easier to come up with three different sets of parameters. The main reason for using a modified set of parameters in the contract games (relative to FS (1999)) was that we wanted to simplify the analysis. Table 1 shows the distributions that we used.

In FKS (2007, p. 144) we say that the second distribution "is a simplification of the distribution that we used in Fehr and Schmidt (1999)". Footnote 17 is added to this sentence in which we report the distribution that we used in FS (1999) and say: "It turns out to be very tedious to solve the model for four different types. This is why we used the simplified model and use only two different types here." Similarly, in FKS (2008, p. 1279) we also call it a "simplification" of the FS (1999) distribution and add a very similar footnote (FN 16) that reports the FS (1999) distribution and explains why we used the simplification.

Why did we use a different distribution at all? There are two reasons. First, the three contract papers consider multi-stage games with incomplete information. Analyzing these games with a four point distribution of types is very tedious indeed. With the two-point distribution we need an Appendix of 10–20 pages in each of these papers to analyze the experimental games under consideration. We did not do it with four types, nor did Binmore and Shaked. However, we did check at several points how robust our results are with respect to the 60:40 assumption.

Second, the purpose of the contract papers is neither to test the theory of inequity aversion nor the four point distribution we suggested there. The section with the theoretical analysis is called "A Theoretical Interpretation of the Results" in all three papers. We very clearly state that these experiments are not suited to test the theory of inequity aversion:

> "It is not our aim here to test the theory of inequity aversion relative to other theories of fairness and reciprocity. Instead we use the theory to acquire a better understanding of the possible mechanisms driving our results." (FKS 2007, p. 144, FN 16).[10]

We explicitly state that "other theories of fairness and reciprocity" may also "be able to rationalize the data" and that we use the model of inequity aversion only because of its tractability. The purpose of the exercise is "to seek an explanation of these puzzles" (the poor performance of trust contracts and the good performance of bonus contracts) "in the context of recently developed fairness models".

The games considered in the contract papers are multi-stage games with incomplete information that are not suitable for the identification of the precise form of other-regarding preferences. There are, however, several other papers that actually test the predictions of inequity aversion, including Kagel and Wolfe (2001), Andreoni and Miller (2002), Charness and Rabin (2002), Falk et al. (2003), Engelmann and Strobel (2004), Fehr et al. (2006), Bolton and Ockenfels (2006), Korenok et al. (2008) and Stanca (2008).

6. The usefulness of incomplete theories

Finally, Binmore and Shaked claim that the model of inequity aversion does not explain the outcome of the contract games even if the simplified two-type distribution is used. They argue as follows: the contract games assume that 60 percent of the subjects are purely self-interested and will take actions that maximize their monetary payoffs, while 40 percent are inequity averse and will take actions that equalize payoffs. If we observe that the ratio of money-maximizing to money equalizing choices differs from the 60:40 assumption at one stage of the game, the theory is refuted and we cannot use it to explain behavior at later stages of the game.

This point raises a serious question: When should a theory of social behavior be considered as refuted? And when should the theory be considered as useful even if not all individual choices are in line with the model? A strict falsificationist would argue as follows: Suppose that a theory predicts that for all individuals i it is true that if i is in situation x then he or she chooses action y. If we conduct an experiment with many individuals in situation x, and if we observe that at least one individual does not choose y but rather z, then the theory is refuted. According to such an extreme standard any theory of social behavior is refuted.

[10] Similar statements can be found in the other two contract papers. See Fehr, Kremhelmer and Schmidt (2008, p. 1277) and Fehr and Schmidt (2004, p. 469).

In our view a model of social behavior is always an idealization that focuses on some forces affecting individual behavior and abstracts from many others. Whether a particular model is a good one depends on the situation under consideration and on the question that is being addressed. Even if most subjects choose actions that differ more or less from the prediction of a particular model, it is still possible that the model generates good predictions at an aggregate level and that it can be used to better understand important forces that are driving behavior in the experiment. For example, neoclassical price theory assumes that all people are fully rational and purely self-interested. We know from many experiments that most people are not fully rational and are not only concerned about their narrow self-interest. Nevertheless, neoclassical price theory is a good model to predict aggregate outcomes in competitive markets and to help us understand how competition affects behavior. As we mentioned in Section 2 above, more general models of social preferences like FS (1999) and Duwfenberg et al. (2008) help us to understand why this is case.[11]

Let us get back to Binmore and Shaked's criticism of the contract papers. We mentioned already in FS (1999) that inequity aversion does not provide a complete description of other-regarding preferences because reciprocity motives may also play a role. Thus, to the extent to which other social motives than inequity aversion are relevant, the theory of inequity aversion cannot provide a full explanation of all individual behaviors in the contract games. There are also other reasons why the theory may fail to explain all individual behaviors. For example, if some subjects have preferences characterized by nonlinear inequality aversion they do not perfectly equalize monetary payoffs but they still move in the direction of more equality. We explicitly abstained from using this more general model because it is less tractable and because it gives more degrees of freedom to fit the model to the data. The cost is, of course, that it is easier to refute the model. Nevertheless, even if the theory does not fit the data perfectly, it may still be useful in helping us understand key features of the data as the following example illustrates.

The experiments in FKS (2007) show that trust contracts, where the principal pays a generous wage upfront, do very poorly and are not profitable, while bonus contracts, where the principal may make a voluntary bonus payment ex post, do very well and are highly profitable. Both contracts appeal to fairness and reciprocity. The trust contract appeals to the agent's fairness to reciprocate a generous wage with a high effort. The bonus contract appeals to the principal's fairness to reciprocate a high effort with a generous bonus. Why does the former so poorly and the latter so well? We establish a general result:

> *Independent of the distribution of fair and selfish types and independent of the specific parameters of the experiment, the theory of inequity aversion predicts that bonus contracts do always better than trust contracts!*

Again we have to refer the reader to FKS (2007, Proposition 3) for an explanation of why this is the case. Our point here is that the theory of inequity aversion helps us to see and understand a general pattern that we did not understand beforehand. This is what we needed the theory for.[12]

Another problem with the argument of Binmore and Shaked is that classifying subjects as "selfish" or "inequity averse" on the basis of the data from the contract experiments is far from trivial. To see this, consider the experiments of FKS (2007). If a principal offers a bonus contract, he has to decide on his voluntary bonus payment at the last stage of the game. Some subjects consistently choose a bonus of zero and can safely be classified as "selfish". Some subjects consistently choose a bonus that equalizes payoffs. They can be classified as inequity averse. However, many subjects choose something in between, and some of them do not behave consistently over time (for the same effort level they sometimes pay a bonus, sometimes they do not).[13] It could be argued that a subject who offers a positive bonus is not selfish, so he or she should be counted as inequity averse. But it could also be argued that a subject who does not equalize payoffs is not inequity averse and should therefore be classified as selfish. How should we classify a subject that behaves inconsistently? Very different distributions are possible, depending on which classification is chosen.

To circumvent this problem we followed a different approach. We compared the average behavior observed at each stage of the game to the average behavior the model predicts. Table 2 shows the results for the bonus contracts in FKS (2007).[14]

We also use OLS and Tobit regressions to estimate how subjects reacted to different contracts, wages, or effort levels and then compare this to the reaction the model predicts. For example, the OLS regression of bonus contracts shows that if the agent increases his effort by one unit, then the average bonus increases by 2.86. The model predicts that if the agent

[11] See Schmidt (2009) for a more thorough discussion of these points.

[12] Other authors also have found our theory useful for gaining general insights into the problem at hand. For example, Kosfeld et al. (2009) examine the endogenous formation of institutions for the provision of public goods. By applying our theory to this problem they are better able to make sense of the observed data patterns. Likewise, Goeree et al. (2009) apply the theory to the endogenous formation of networks which enables them to understand otherwise puzzling empirical observations.

[13] We acknowledge this in FKS (2007, p. 140): "Although the principals respond, on average, quite strongly to increases in the effort level, it is important to notice that there are big differences in individual behavior. This is reported in Table V." (Table V reports the distribution of bonus payments for any given effort level.) On page 148, we add: "However, it also has to be said that there is a lot of noise in the data and that the model does less well in predicting individual behavior. There are some principals whose behavior is consistent with the model and who either choose $b = 0$ in all periods or a bonus that (roughly) equalizes payoffs. However, there are also principals who pay a positive but smaller bonus and many of them do not behave consistently over time".

[14] In Fehr and Schmidt (2004) and Fehr et al. (2008) we follow the same approach. We refer the reader to the detailed results given there.

Table 2
Predicted average behavior and observed average behavior in the contract game of Fehr et al. (2007).

	Predicted average behavior	Observed average behavior
Stage 3	Predicted average bonus: 6.2	Observed average bonus: 10.4
Stage 2	Predicted average effort: 5.4	Observed average effort: 5.2
Stage 1	Predicted average wage: 15.0	Observed average wage: 15.0
	Predicted fraction of bonus contracts: 100%	Observed fraction of bonus contracts: 87.8%

increases his effort by one unit, the average bonus increase is between 2.2 and 2.8 (depending on the marginal cost of effort). We refer the reader to FKS (2007, including the Appendix) for details.

We do not claim that the predictions of the model of inequity aversion are always as quantitatively accurate as they are in Table 2. Nor do we claim that that these results are evidence in favor of the existence of inequity aversion as a motive. It is important to keep in mind that the contract experiments were not designed to test the theory of inequity aversion. In order to test the motivational assumptions of the theory different types of experiments are needed that can discriminate between different behavioral hypotheses. The purpose of the contract experiments was to study how subjects deal with contractual problems in the lab. We used the theory of inequity aversion to interpret and better understand the observed data patterns such as the clear superiority of the bonus contract over the trust contract.

7. Conclusions

The preceding sections have shown that the criticisms of Binmore and Shaked are not substantiated. We hope that this settles an unproductive debate and frees resources for solving the really important questions. Although there has been much progress in the research on social behavior and other-regarding preferences over the last 15 years experimental and behavioral economists are aware of the limitations of current social preference models. Like others (Andreoni et al., 2003; Andreoni and Miller, 2002; Bellemare et al., 2008a,b; Bolton et al., 1998a,b; Brandts and Charness, 2003; Brandts and Sola, 2001; Charness and Rabin, 2002; Cox, 2004; Dawes et al., 2007; Engelmann and Strobel, 2004; Kagel and Wolfe, 2001; Korenok et al., 2008; Stanca, 2008), we have been actively involved in the conduct of experiments (Falk et al., 2003; Falk et al., 2008; Fehr et al., 2007) that enable us to identify the different types of social preferences and the limits of the current models. The motivational forces assumed in our model are of limited importance in some of the above-mentioned experiments, while a substantial share of subjects seems to make inequity averse choices in others. Furthermore, these experiments indicate the importance of other forms of social preferences, such as the desire to increase the group's overall welfare (Andreoni and Miller, 2002; Charness and Rabin, 2002; Engelmann and Strobel, 2004), the desire to be viewed by (anonymous) others as altruistic (Dana et al., 2006; Ellingsen and Johannesson, 2007), or the desire to reciprocate in response to kind or hostile acts (Falk et al., 2008; Offerman, 2002).

At the current state of empirical knowledge about social preferences the following two tasks are of prime importance. First, we need a parsimonious empirical characterization of the distribution of social preference types. Recent applications of the finite mixture approach in the domain of risk taking provide a rigorous and parsimonious characterization. For example, the papers by Bruhin et al. (2007) and Conte et al. (in press) show that roughly 20% of the population across several different data sets – two Swiss, one Chinese and one British – can be characterized as expected utility maximizers, while 80% are classified as prospect theory types because they exhibit nonlinear probability weighing. The application of the finite mixture approach to the domain of social preferences might achieve a similarly parsimonious characterization of social preference types.

Second, theorists will have to build models capable of simultaneously incorporating different types of social preferences on the basis of the empirical results. However, there is no guarantee that the outcome of this endeavor will provide us with an empirically informed model that is tractable enough for applied economics. Thus, we may be faced with a situation where we have to sacrifice precision in the modeling of individual preferences in order to keep the models tractable for applied work. For this reason, we believe that simple models – such as the theory of inequity aversion or the consequentialist part of Charness and Rabin's theory of other-regarding preferences – that deliberately abstract away from some empirically observed facts in order to maintain tractability – will continue to play a role.

Both of these models are clearly simplistic. They assume a linear functional form and maintain that all actors are perfectly rational. Clearly, this does not reflect the complexity of human behavior. Nevertheless, we believe that such theories can be very useful. They are a tool that helps us think in a systematic and logical way about what we observe, they offer a new perspective on the data, they yield specific predictions that can be tested, and they point to interesting new experiments and theoretical hypotheses.

References

Andreoni, I., 1995. Cooperation in public-goods experiments: kindness or confusion? American Economic Review 85 (4), 891–904.
Andreoni, I., 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., 2002. Giving according to garp: an experimental test of the consistenca of preferences for altruism. Econometrica 70 (2), 737–753.

Bellemare, C., Kroger, S., van Soest, A., 2008a. Preferences, intentions, and expectations: a large-scale experiment with a representative subject pool. IZA Working paper No. 3022.
Bellemare, C., Kroger, S., van Soest, A., 2008b. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. Econometrica 76 (4), 815–839.
Binmore, K., Shaked, A., 2010. Experimental economics: where next? Journal of Economic Behavior & Organization 73 (1), 120–121.
Bolton, G., Brandts, J., Ockenfels, A., 1998a. Measuring motivations for the reciprocal responses observed in a simple dilemma game. Experimental Economics 1 (3), 207–219.
Bolton, G., Katok, E., Zwick, R., 1998b. Dictator game giving: rules of fairness versus acts of kindness. International Journal of Game Theory 27 (2), 269–299.
Bolton, G., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. American Economic Review 90 (1), 166–193.
Bolton, G., Ockenfels, A., 2006. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment. American Economic Review 96 (5), 1906–1911.
Brandts, J., Charness, G., 2003. Truth or consequences: an experiment. Management Science 49 (1), 116–130.
Brandts, J., Sola, C., 2001. Reference points and negative reciprocity in simple sequential games. Games and Economic Behavior 36, 138–157.
Bruhin, A., Fehr-Duda, H., Epper, T., 2007. Risk and rationality: uncovering heterogeneity in probability distortion. Working Paper 0705, University of Zurich, Socioeconomic Institute.
Camerer, C., 2003. Behavioral Game Theory—Experiments in Strategic Interaction. Princeton University Press, Princeton, New Jersey.
Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Quarterly Journal of Economics 117 (3), 817–869.
Conte, A., Hey, J., Moffat, P., in press. Mixture models of choice under risk. Journal of Econometrics.
Cox, J., 2004. How to identify trust and reciprocity. Games and Economic Behavior 46 (2), 260–281.
Dana, J., Cain, D., Dawes, R., 2006. What you don't know won't hurt me: costly (but quiet) exit in dictator games. Organizational Behavior and Human Decision Processes 100 (2), 193–201.
Dawes, C., Fowler, J., Johnson, T., McElreath, R., Smirnov, O., 2007. Egalitarian motives in humans. Nature 446 (7137), 794–796.
Dufwenberg, M., Heidues, P., Kirchsteiger, G., Riedel, F., Sobel, J., 2008. Other regarding preferences in general equilibrium. CEPR Discussion Paper No. 6815.
Eckel, C, Gintis, H., in press. The state of experimental economics: comments on Binmore and Shaked. Journal of Economic Behavior & Organization.
Ellingsen, T., Johannesson, M., 2007. Paying respect. Journal of Economic Perspectives 21 (4), 135–149.
Engelmann, D., Strobel, M., 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. American Economic Review 94 (4), 857–869.
Falk, A., Fehr, E., Fischbacher, U., 2003. On the nature of fair behavior. Economic Inquiry 41 (1), 20–26.
Falk, A., Fehr, E., Fischbacher, U., 2008. Testing theories of fairness—intentions matter. Games and Economic Behavior 62 (1), 287–303.
Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games and Economic Behavior 54 (2), 293–315.
Fehr, E., Kremhelmer, S., Schmidt, K., 2008. Fairness and the optimal allocation of ownership rights. Economic Journal 118 (531), 1262–1284.
Fehr, E., Schmidt, K., 2006. The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories. In: Kolm, S., Ythier, J. (Eds.), Handbook on the Economics of Giving, Reciprocity and Altruism, Vol.1. Elsevier, Amsterdam, pp. 615–669.
Fehr, E., Naef, M., Schmidt, K., 2006. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment. American Economic Review 96 (5), 1912–1917.
Fehr, E., Schmidt, K., 2003. Theories of fairness and reciprocity: evidence and economic applications. In: Dewatripont, M., Hansen, L., Turnovky, S. (Eds.), Advances in Economics and Econometrics. Cambridge University Press, Cambridge, pp. 208–257.
Fehr, E., Schmidt, K., 2004. Fairness and incentives in a multi-task principal-agent model. Scandinavian Journal of Economics 106 (3), 453–474.
Fehr, E., Schmidt, K., 2005. The Rhetoric of Inequity Aversion—A Reply. University of Munich, Mimeo.
Fehr, E., Klein, A., Schmidt, K., 2007. Fairness and contract design. Econometrica 75 (1), 121–154.
Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics 114 (3), 817–868.
Fischbacher, U., Gachter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters 71 (3), 397–404.
Goeree, J., Riedl, A., Ule, A., 2009. In search of stars: network formation among heterogeneous agents. Games and Economic Behavior 67 (2), 445–466.
Kagel, J., Wolfe, K., 2001. Tests of fairness models based on equity considerations in a three-person ultimatum game. Experimental Economics 4 (3), 203–219.
Kocher, M., Cherry, T., Kroll, S., Netzer, R., Sutter, M., 2008. Conditional cooperation on three continents. Economics Letters 101, 175–178.
Korenok, O., Millner, E., Razzolini, L., 2008. Experimental evidence on inequality aversion: dictators give to help the less fortunate. Working Paper, Virginia Commonwealth University.
Kosfeld, M., Okada, A., Riedl, A., 2009. Institution Formation in Public Goods Games. American Economic Review 99 (4), 1335–1355.
Levine, D., 1998. Modeling altruism and spitefulness in experiments. Review of Economic Dynamics 1 (3), 593–622.
Offerman, T., 2002. Hurting hurts more than helping helps. European Economic Review 46 (8), 1423–1437.
Roth, A., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburg and Tokyo: an experimental study. American Economic Review 81 (5), 1068–1095.
Schmidt, K., 2009. The role of experiments for the development of economic theories. Perspektiven der Wirtschaftspolitik 10 (s1), 14–30.
Shaked, A., 2005. The Rhetoric of Inequity Aversion. University of Bonn, Mimeo.
Stanca, L., 2008. How to be kind? Outcomes versus intentions as determinants of fairness. Working Paper No. 145, Department of Economics, University of Milan, Bicocca.