# A Flawed Case for Inequity Aversion: Revisiting the Criticisms of Binmore and Shaked (2010)

Ammar Plumber

## Introduction

Ernst Fehr and Klaus Schmidt advocate the use of an "inequity aversion" utility function in game theoretic modeling as a means of explaining and predicting experimental data that money-maximizing models fail to accommodate. In response to their work, two prominent critical papers have been put forth: Shaked (2005) and Binmore and Shaked (2010). This paper aims to explicate and evaluate the criticisms leveled against Fehr and Schmidt in the second paper.

First, I describe the theory of inequity aversion: how it is formalized, calibrated, and tested. Furthermore, I will identify exactly what data Fehr and Schmidt are seeking to explain in proposing the theory and how they think that their model outperforms the available alternatives. Then, I discuss a few thematic criticisms. One concerns the communicative tactics that Fehr and Schmidt use to mask deficiencies of their theory. Another question is whether their theory is predictively successful in the first place—especially with regard to the four games that Fehr and Schmidt consider. Finally, Binmore and Shaked (2010) allege that Fehr and Schmidt use different parametric distributions to fit the data, a tactic that resists falsification. In adjudicating these criticisms, I consider the responses given in Fehr and Schmidt (2010). To conclude, I discuss how these criticisms weigh on the credibility of Fehr and Schmidt's model.

## The Inequity Aversion Model: An Overview

Fehr and Schmidt declare that their intention in proposing the theory of inequity aversion was to answer the following question: why do subjects act selfishly in some games but generously and fairly in others? The simple theory that they settled on was one that they found to be "a tractable tool in more complicated models and that yields quantitative, testable predictions" (p 101).

To clarify the need for an alternative to the standard money-maximizing model, Fehr and Schmidt (1999, pp 817-818) point to a few experimental findings. A number of bilateral bargaining experiments seem to indicate that not all subjects are purely money-maximizers and in fact value fairness (Güth, Schmittberger, and Tietz 1990; Roth 1995; Camerer and Thaler

1995). In other contexts, however, it seems that fairness considerations are not particularly salient (Smith and Williams, 1990; Roth, Prasnikar, Okuno-Fujiwara, and Zamir, 1991; Kachelmeier and Shehata, 1992; Güth, Marchand, and Rulliere, 1997). By Fehr and Schmidt's assessment, traditional money-maximizing models do not explain these varied findings well, so they apply the assumption that a fraction of people are motivated by fairness—modeled as self-centered inequity aversion. This assumption is formalized according to the following utility function:

$$U_i(x) = x_i - \frac{\alpha_i}{n-1}\sum_{j \neq i}(x_j - x_i)^+ - \frac{\beta_i}{n-1}\sum_{j \neq i}(x_i - x_j)^+$$

where $0 \leq \beta_i < 1$, $\beta_i \leq \alpha_i$ and $x^+ = \max\{x, 0\}$. The parameter $\alpha$ represents the strength of one's aversion to others receiving more than oneself, and $\beta$ is the strength of one's aversion to receiving more than others. Because the signs of the last two terms are negative, greater values of $\alpha$ and $\beta$ are associated with utility reductions stemming from inequities.

The inequity aversion theory's *prima facie* domain of application extends to all games with more than one player. Fehr and Schmidt place no apparent constraint on where their theory should be applied besides this self-evident multi-subject requirement. However, Binmore and Shaked (2010, p 91) rightly point out that context bears heavily on what people count as fair. Two specific factors affecting these evaluations include "perceived need and prior investment of effort" (p 91). These contextual factors place an implicit constraint on the theory's domain of application; insofar as assessments of fairness are not straightforward, it is unclear how Fehr and Schmidt's utility function can be applied. Thus, Binmore and Shaked (2010) restrict their attention to the applications that Fehr and Schmidt propose with apparent success.

Another question concerns *how* the theory is applied given Fehr and Schmidt's assumption that people vary in their degrees of inequity aversion. In other words, the values of $\alpha$ and $\beta$ are heterogenous across a population. Note that predictions can only be derived from the inequity aversion model if the values of $\alpha$ and $\beta$ are known. But, every experiment or situation potentially involves a different sample, and there is no value of $\alpha$ and $\beta$ that is universally held. The most straightforward solution would be to identify a marginal and joint distribution of $\alpha$ and $\beta$, and one can assume that each experimental sample is representative of the population distribution. Fehr and Schmidt (1999, 2003) do exactly this using experimental data from the ultimatum game, defined in the forthcoming section. Binmore and Shaked (2010, p 92) question whether Fehr and Schmidt keep these distributions constant when predicting data in future experiments, a criticism that I shall return to in Section III.

Finally, to apprehend the stakes of these criticisms, one should note the popularity of the inequity aversion model. The landmark *QJE* paper by Fehr and Schmidt (1999) has been cited 11,396 times according to Google Scholar—a six-fold increase since 2010. Binmore and Shaked (2010, p 90) examine the breakdown of these papers and find that few question the empirical adequacy of inequity aversion or its appropriate range of application. Most cite their work as an established development in the field of behavioral economics. Some even consider the model important enough to be worth teaching to undergraduates (Wilkinson, 2008). It must either be the case that Binmore and Shaked's criticisms have been overlooked by the academic community or have been duly assessed as unproblematic. I consider the latter case to be unlikely.

## Games, Experimental Designs, and Other Terms

In the forthcoming sections, I will discuss and allude to four different types of games to which Fehr and Schmidt apply their model: the ultimatum game, the public goods game, auction games, and bonus contract games.

The ultimatum game involves two parties, anonymous to one another. The proposer, endowed with a sum of money, proposes a certain division to a responder, who may either accept or veto the deal. If the responder accepts, the two players receive money as per the agreed upon division. If the responder rejects the offer, neither player receives money. Interestingly, a breadth of experimental data has shown that responders in a one-shot game consistently reject low offers and prefer receiving nothing to a disproportionately small monetary gain. Fehr and Schmidt (1999) point to this finding as a phenomenon that the inequity aversion model helps to explain; fairness and reciprocity appear to be salient concerns for ultimatum game subjects. Otherwise, Fehr and Schmidt argue, all gainful offers would be accepted.

In the public goods game, subjects are allotted individual budgets. From these, subjects can secretly make donations to be put into a shared pot, which is multiplied at each stage by a growth factor and divided evenly between players. Subjects may retain the resources that they do not donate. Players who opt not to donate but receive a share of the pot are referred to as "free riders". "Cooperators", on the other hand, contribute an average or above average amount to the shared pot. There are a few design specifications that may influence experimental results. First, subjects may be given the option to punish free riders by eliminating their payoffs. This specification is referred to as the public goods game with punishment. Note that in the experiments studied by Fehr and Schmidt (1999) punishment is costly for those who punish. Public goods games can either proceed in multiple stages or in one. When the same subjects play the public goods game for a series of rounds, this is referred to as an iterated public goods game or a partner design. When subjects only interact with one another for one round, this is referred to as a one-shot public goods game or a stranger design. The stranger design is the standard experimental technique because the partner design introduces the variable of shared experience, which complicates how data is to be interpreted. As I will discuss, Fehr and Schmidt interestingly opt to use the partner design. In any case, they consider the public goods game to be a crucial test of the theory, as they assert that experimental findings diverge from the predictions of the money-maximizing model. Thus, Fehr and Schmidt argue, their theory may better explain the data.

There are two varieties of auction or market games that Fehr and Schmidt consider. The first is an auction game with proposer competition, which proceeds as follows. Multiple proposers are each afforded a sum of money. All proposers offer a share of their budget to a single responder, who will accept or reject the highest offer. If a proposer's offer is rejected, neither does the responder receive any money nor may the proposer keep any of his/her initial funds. If one proposer makes the highest bid and if the bid is accepted, the money is divided accordingly. If multiple proposers are tied for the highest accepted bid, one is randomly chosen to divide his/her budget with the responder. The second kind of auction game that Fehr and Schmidt examine involves responder competition. This scenario involves a single proposer making a uniform offer to many responders. Among accepting responders, one is chosen at random to divide the budget with the proposer. Fehr and Schmidt admit that the money-maximizing model performs well with market games. Their aim in examining it is only to show that the inequity aversion model performs equally well.

In a number of papers, Fehr and Schmidt different kinds of contract games. For the purposes of this paper, I focus on bonus contract games. The experiments in question involve principals giving contracts to an agent. The principal stipulates a wage, an effort level, and a bonus. Neither the agent's effort nor the bonus are contractually enforceable. At a final stage, the principal chooses whether or not to award the bonus. Fehr and Schmidt (2005) point to the data generated from bonus contract experiments as consistent with the aggregate predictions of the inequity aversion model. One reason Fehr, Klein, and Schmidt (2007) think it important to examine bonus contracts is that these contracts contain an implicit fairness clause; if agents expend a satisfactory amount of effort, it is fair that a bonus would be awarded. Thus, subjects' behavior in stipulating and awarding bonus contracts is perhaps a crucial test of the inequity aversion model.

**Favorable Selection: A Matter of Academic Integrity**

The first major charge leveled against Fehr and Schmidt is that they fail to disclose a number of tactics that portray their theory more favorably. The problematic tactics in question are examples of what Binmore and Shaked (2010) refer to as "favorable selection". I consider three cases. First, Fehr and Schmidt (1999) count auction games as a predictive success when many other models can predict the same events with equivalent success. Second, for the public goods game without punishment, Fehr and Schmidt (1999) opt for the partner design rather than the standard stranger design, which produces starkly different outcomes than what the inequity aversion model would predict. In addition, they favorably select the rival theory against which they evaluate their own model's performance. Third, in analyzing bonus contract games (Fehr and Schmidt, 2004; Fehr et al, 2005; Fehr et al, 2007), Fehr and Schmidt selectively ignore the many failed predictions antecedent to the averages they find to be in line with their model's predictions. I shall discuss each of these criticisms in turn and assess their veracity.

If data from some experiment is predicted successfully by a wide range of rival theories, the same data cannot serve as a challenging test of the adequacy of a new theory, as the events in question are easy to predict. But, Fehr and Schmidt (1999) seem to commit this exact folly by using their model to predict auction games with proposer and responder competition—games that many other models predict with apparent success. Interestingly, under both game specifications, even when traders are assumed to have "zero-intelligence", or bid at random subject to a budget constraint, predicted outcomes are largely in line with empirical data (Binmore and Shaked, 2010, p 95; Gode and Sunder, 1995).

For the auction game with proposer competition, Fehr and Schmidt's model successfully predicts that experienced proposers will offer all surplus to responders. However, they make a mistake in experimental design by failing to allow responders to choose an equitable offer over the highest offer (Binmore and Shaked, 2010, p 95). Had they allowed for this choice, their theory could have been crucially tested—confirmed or disconfirmed. Instead, what they found tells us very little beyond that their model can predict outcomes that are just as easily predicted by other models. For the auction game with responder competition, the inequity aversion model, consistent with observed data, predicts that experienced responders will accept very low offers. Again, the money-maximizing model generates the same prediction. For these reasons, as Binmore and Shaked (2010) declare, this easily predictable game offers little in the way of supporting the inequity aversion theory, and Fehr and Schmidt's decision to feature it in defense of their model is puzzling and perhaps somewhat deceptive.

A second example of favorable selection lies in Fehr and Schmidt's (1999) decision to examine the public goods game without punishment under the partner design specification rather than the stranger design. Subjects tend to play more cooperatively under the partner design, a factor that would inflate the empirical adequacy of an inequity aversion model relative to the stranger design. Because of the dubious empirical adequacy of backward induction, standard practice is to examine one-shot games using the stranger design, which makes it even more peculiar that Fehr and Schmidt would opt for the partner design. Binmore and Shaked (2010, p 93) show that had Fehr and Schmidt examined data from different specifications of the game, their model would have performed substantially worse than it did. Moreover, they do not compare their predictions with the no-punishment experimental data that appears in Fehr and Gächter (2000) despite citing data from this paper when examining the game *with* punishment. As Binmore and Shaked (2010, p 94) note, had they used the numbers from Fehr and Gächter (2000), they would have found an even larger discrepancy between the prediction and data than they did—an indication that the model fails to predict the behavior of nearly 49% of subjects.

When analyzing the public goods game with punishment, Fehr and Schmidt also opt for the partner design. For this variation of the public goods game, it may be argued that the inequity aversion model predicts better than does the money-maximizing model. However, Binmore and Shaked point out three different kinds of favorable selection that Fehr and Schmidt employ in making this claim. First, their model's prediction is more accurate under the partner design specification, wherein rates of maximum contribution are 80%, than under the stranger design specification, wherein only 10% contribute the maximum. Binmore and Shaked note that it is unnecessary for Fehr and Schmidt to use the unconventional partner design, as even the 10% maximum contribution rate of the stranger design is sufficient to challenge the supposedly unique money-maximizing prediction that all subjects will free ride. About this instance of favorable selection, Fehr and Schmidt (2010) say nothing other than to assert that the lower contribution rates of a stranger design would not be inconsistent with their model, which allows for a range of equilibria. However, if any level of contribution can be defended as a prediction of the model, it makes little sense to claim that their model meaningfully or uniquely predicts the data observed under the partner design; their model does not predictively favor the observed equilibrium more than it does any other equilibrium.

Moreover, Fehr and Schmidt favorably select the variation of the money-maximizing model against which they compare their model's performance. For example, money-maximizing models that attribute irrational behavior to a small fraction of the population generate different predictions than what Fehr and Schmidt (1999) regard as the unique money-maximizing prediction (Binmore and Shaked, 2010, p 94; Kreps et al, 1982). In fact, there are a range of Nash equilibria under the money-maximizing model including all players contributing the maximum. This is because even experienced players may have never learned to evaluate subgames that have never been reached. Consider the following example given by Binmore and Shaked (2010, p 95). Assume that every player at the outset intends to punish others who fail to contribute the maximum. Nobody deviates for fear of being punished, and, for this reason, nobody learns that punishing players who will never again be encountered is pointless. As such, all players contributing the maximum is a Nash equilibrium, and Fehr and Schmidt are therefore wrong to uniquely identify the money-maximizing model with the prediction that all players will free ride. In their reply to Binmore and Shaked (2010), Fehr and Schmidt (2010) offer no defense of this alleged instance of favorable selection.

A final instance of favorable selection is evident in Fehr and Schmidt's application of their model to bonus contract games. As previously discussed, the experiments in question involves a principal offering a contract (wage, effort level, and bonus) to an agent. The agent may decide how much effort to exert, and the principal may decide whether or not to award the bonus. Fehr and Schmidt claim to successfully predict the average wage offered by principals, average bonus awarded, and average effort level. Problematically for Fehr and Schmidt, Binmore and Shaked (2010) identify a number of failed predictions antecedent to these predicted averages. For instance, the 40-60 distribution that Fehr and Schmidt assume in order to obtain these averages are clearly inconsistent with the observed percentage of principals who exhibit inequity averse behavior. Moreover, the inequity aversion model would predict that the bonuses offered by inequity-averse principals would equalize payoffs of principals and agents, but the data clearly refutes this prediction. Fehr and Schmidt (2004) and Fehr et al (2005, 2007) ignore these disconfirming observations and conveniently focus instead on the averages. Unquestionably, this is an instance of favorable selection; their model fails to predict all granular empirical findings, of which the averages that their model predicts are derivative. When discussing the inequity aversion model's predictive shortcomings, I further consider these contract games in the next section.

## Predictive Shortcomings of the Inequity Aversion Model

Even when overlooking Fehr and Schmidt's misleading communicative tactics, it appears that the inequity aversion model predicts quite poorly for the games Fehr and Schmidt claim as confirmation of their theory. Binmore and Shaked (2010) substantiate this claim for three kinds of games: public goods games without punishment (partner design), auctioning games, and bonus contract games.

For public goods games without punishment, the inequity aversion model predicts that nearly 98.48% of subjects will be free riders, but the data yields an average of 73%—a relative difference of 35%. Glossing over this considerable difference, Fehr and Schmidt (1999) say, "Thus, it seems fair to say that our model is consistent with the bulk of individual choices in this game" (p 845). Further, the money-maximizing model appears to issue a nearly identical prediction of 100% free riders. Assessing the predictive efficacy of the inequity aversion model requires a standard of comparison: how well the model performs relative to its alternatives. Given that the model hardly performs better with regard to this game than does the money-maximizing model, it makes little sense to count this data as evidence for the theory. Fehr and Schmidt (2010) take a concessionary tone in their response; they emphasize that they claimed only to roughly approximate the choices of experimental subjects. Even if Fehr and Schmidt's defense is true, Binmore and Shaked (2010) are correct to claim that the theory's predictive shortcomings cast doubt on its adequacy for explanatory purposes. After all, a theory cannot be said to make sense of the data—as Fehr and Schmidt assert—if it entails outcomes that are empirically contradicted. I return later to a discussion of the related aims of prediction and explanation.

As previously discussed, the same appears to be true for the auctioning games, as many alternative models predict the same outcomes; even if players bid at random subject to a budget constraint, predictions are consistent with experimental data (Gode and Sunder, 1995). Money-maximizing models, in particular, do very well with these games. Thus, auctioning games offer no predictive basis for preferring the inequity aversion model, and in a footnote, Fehr and Schmidt (2010) admit as much (p 102).

Finally, the inequity aversion model fails to predict bonus contract games in a number of ways. Fehr and Schmidt assume for their predictive purposes that 40% of individuals are inequity averse. For the sample population in Fehr et al (2007), it seems that only 27% of principals are inequity averse. If Fehr and Schmidt were to use a 27-73 ratio of inequity averse individuals to money-maximizers, the theory predicts no cooperation between principals and agents, with money-maximizing agents choosing low effort levels and principals generally opting not to award bonuses—a far cry from what the 40-60 distribution predicts. The same appears to be true of the fraction of inequity-averse individuals observed in Fehr and Schmidt (2004); 15.7% of individuals exhibit behavior consistent with inequity aversion, and the 40-60 distribution would contradict the non-cooperative consequent. Even if Fehr and Schmidt successfully predict the average bonus, average initial offer, and average effort level, they fail to predict any of the decisions and distributions that these averages stem from—hence, the charge of favorable selection by Binmore and Shaked (2010). What this shows is that there must exist some rival theory that would accurately predict granular behavior as well as aggregate tendencies, and, as such, the data seems to disconfirm inequity aversion theory rather than supporting it.

To understand the implications of these predictive deficiencies, a couple of questions must be explored. First, to what extent is behavioral game theory a predictive endeavor, an explanatory one, or both? With these aims in mind, why might one prefer the inequity aversion model to its alternatives? Finally, do the inequity aversion model's predictive and explanatory features bear on its optimal domain of application? I explore each of these questions in turn.

On the explanatory account of game theory, one would regard game theory as accounting for the *reasoning* processes that motivate behavioral outcomes. The defense offered by Fehr and Schmidt (2010) seems to echo this claim. An explanatory conception of game theory asks *why* subjects act in certain ways. For Fehr and Schmidt, the salient question is why experimental data reveals that subjects sometimes act as if they are not pure money-maximizers and perhaps exhibit other-regarding preferences (Fehr and Schmidt, 2010, p 102-103). The inequity aversion model was put forth in an effort to explain this finding. It assumes that subjects act out an ingrained psychological preference for more equal outcomes.

A predictive account of game theory, on the other hand, focuses on mere strategic *behavior*, not reasoning. *Why* subjects act as they do is irrelevant (and often unknowable). All that is relevant is that *how* they act is consistent with what the model predicts. Whether Fehr and Schmidt's work can be defended depends largely on how one understands the game theoretic task *qua* prediction or explanation. I examine how the inequity aversion model is assessed and applied in order to determine whether prediction or explanation is more fundamental. The inequity aversion model is useful for explanatory purposes if experimental data reveals subjects behaving *as if* they are averse to inequities. Whether or not inequities factor into a subject's motivations is not observable from behavior alone. Thus, to the extent that behavior deviates from the model's predictions, the motivations that the model attributes to subjects cannot suffice for explanatory purposes. Predictive deficiencies, therefore, weigh heavily against the explanatory virtues that Fehr and Schmidt claim in support of their theory. Second, one could question whether game theoretic modeling serves an explanatory purpose at all. Game theory seeks to model social behavior, not motivations.

Hypotheses about motivations, however, may emerge as a byproduct of predictive successes and failures, and psychological studies may follow. Whether game theory plays an explanatory role is beyond the scope of this paper, but suffice it to say that the predictive shortcomings of the inequity aversion model certainly cast doubt on its explanatory value. There are multiple reasons why one model might be preferable to another. A few include range of application or universality, ability to account for anomalous data, simplicity or elegance, and consistency with wider bodies of knowledge (psychological theory, for instance).

A more universal theory can be applied to a wider range of situations than can a more provincial one. Accordingly, Fehr and Schmid (1999) claim that their theory "can be applied to any game" (p 856). However, as Binmore and Shaked (2010) point out, what subjects perceive as fair and the extent to which they value fairness depends on a number of contextual situations; people do not expect others to evenly split the contents of their wallets with strangers they pass on the street even if they may prefer to do so in an ultimatum game setup. On the other hand, the social norms approach appears to be more universally applicable relative to the inequity aversion model. Social norms are possibly present in all social contexts, and, once identified, subjects' sensitivity to the norm can be measured empirically. A relevant consideration in assessing a theory's proper domain of application is in which contexts it predicts better than rival theories. For the inequity aversion model to have substantial appeal, there must be a reasonable number of contexts in which it performs better than its alternatives.

It seems, as per the predictive deficiencies highlighted by Binmore and Shaked (2010), that the model does not do so.

Fehr and Schmidt claim as an advantage of their theory that it predicts experimental findings that money-maximizing models cannot account for—what one might call anomalous data, unassimilable by the dominant paradigm. The puzzle in question that Fehr and Schmidt claim to solve is why other-regarding preferences appear to be operative in some subjects and contexts but not in others. The favorable selection problem I discuss earlier impugns this claim, as much of the data that they claim as anomalous would in fact be accommodated by reasonable variations of the money-maximizing model. Moreover, the predictions that do appear more accurate are not unique predictions of the inequity aversion model; a number of less accurate predictions could be claimed as equilibria. Finally, as this section recounts, the model fails to predict several games that Fehr and Schmidt consider to be crucial tests of their theory. For these reasons, I conclude that the inequity aversion model does not deal with anomalous findings significantly better than do other models.

Another purported virtue of the inequity aversion model is that it is simple and tractable. In other words, it assumes relatively little and can be applied to a variety of situations with few formal adjustments. Tractability is hardly an advantage of this theory over its major rivals—the social norms and money-maximizing paradigms. Each of these alternatives can be just as easily applied to different situations, and the social norms approach, in particular, offers the ability to directly measure the importance of a norm rather than calibrating a parameter using a different experimental situation, as in the case of the inequity aversion model. It is true that the inequity aversion model is simple, however, but, again, no simpler than its rivals. Both the social norms and money-maximizing approaches posit very little about subjects' motivations. In fact, one could argue that the inequity aversion model assumes more by assuming from the outset that fairness considerations are present. The social norms approach, on the other hand, measures the importance of a norm rather than assuming it. Neither simplicity not tractability, it seems, are relative strengths of Fehr and Schmidt's model.

A model's consistency with theory in other domains may also be claimed as an advantage over its rivals. If an empirically adequate model expresses certain synergies with successful theories in psychology, for instance, it may be preferable to models whose assumptions contradict those same theories. For instance, money-maximizing models fail to account for psychological motives that often yield non-money-maximizing outcomes. This theoretical virtue to some extent reflects the explanatory aims of theory, though the explanatory relevance of game theoretic models comes not from the model alone but from its ability to successfully predict using realistic assumptions. On this view, one must also hold that realism is a crucial feature of behavioral models. Fehr and Schmidt (1999) claim this advantage for their theory and say explicitly, "Our theory is motivated by the psychological evidence on social comparison and loss aversion" (p 856). Insofar as the psychology they appeal to seems conclusive, this is a virtue of the inequity aversion model. However, Binmore and Shaked (2010) prefer the social norms approach on the grounds that subjects' other-regarding preferences are context-dependent. Thus, the psychological theories that Fehr and Schmidt apply are only conditionally applicable, and, in behavioral experiments, one can never know *a priori* which conscious or subconscious motives are operative. A social norms approach perhaps better account for these contextual considerations that are only measurable empirically, *a posteriori*. The inequity aversion model, therefore, while formalizing established psychological motives, does not do so in a manner that is comprehensive and is not justified in its generalizing assumption.

## Methodological Concerns: Setting Parameters

Fehr and Schmidt's use of parameters poses an additional methodological concern. In assessing a theory, it is necessary to compare forecasts to data, and the inequity aversion model is formalized such that parameters must be set before forecasts can be produced. Much of the forecasting successes that Fehr and Schmidt tout in defense their theory rely on a parametric distribution abstracted from ultimatum game experimental data. Binmore and Shaked (2010) allege that Fehr and Schmidt use floating parameters; they revise the distributions they use according to the data that they wish to predict.

Binmore and Shaked focus on a few specific cases of Fehr and Schmidt deviating from their "calibrated" distributions. The calibration they settle upon in their *Quarterly Journal of Economics* paper (1999) is as follows:

| $\alpha$ | % |
|---|---|
| 0.0 | 30 |
| 0.5 | 30 |
| 1.0 | 30 |
| 4.00 | 10 |

marginal distribution

| $\beta$ | % |
|---|---|
| 0.0 | 30 |
| 0.25 | 30 |
| 0.6 | 40 |

marginal distribution

| $(\alpha, \beta)$ | % |
|---|---|
| (0.0,0.0) | 30 |
| (0.5,0.25) | 30 |
| (1.0,0.6) | 30 |
| (4.0,0.6) | 10 |

joint distribution

The four types of individuals are shown in the rightmost table. Each joint distribution corresponds to a different degree of inequity aversion. While Fehr and Schmidt (1999) make clear that these distributions reflect their ultimatum game data, it is less clear exactly *how* they do so. For instance, as Binmore and Shaked (2010, p 92) note, no data is given pertaining to the joint distribution of $\alpha$ and $\beta$. This observation suggests that the parameters are under-identified and may float to a considerable degree. The fact that these parameters can float is evidenced in Fehr and Schmidt's analysis of three contract games (Binmore and Shaked, 2010, pp 96-98). Three papers exhibit an altered distribution from the one declared in the *QJE* paper (1999): one on bonus, trust and incentive contracts (Fehr at al, 2007); another on joint ownership contracts (Fehr et al, 2005); and a final one on piecewise and bonus contracts (Fehr and Schmidt, 2004). In each, Fehr and Schmidt use a 40-60 distribution, wherein 40% of subjects are inequity averse ($\alpha$ = 2 and $\beta$ = 0.6) and 60% of subjects are money-maximizers, with $\alpha$ and $\beta$ values of zero. This distribution treats as money-maximizers the 30% of subjects in the *QJE* joint distribution for whom $\alpha$ = 0.5 and $\beta$ = 0.25. Binmore and Shaked (2010, p 97) correctly note that this distributional difference is meaningful; subjects with mild inequity aversion are not behaviorally identical to pure money-maximizers. Thus, it seems that the choice to deviate from the *QJE* distribution must be justified.

Fehr and Schmidt (2010, p 105) defend this decision by arguing that the 40-60 distribution is a simplification of the distribution used in Fehr and Schmidt (1999), as it would be too tedious and inelegant to consider four types of individuals in their analyses of contract games. As previously noted, calling the difference a mere simplification does not suffice; reclassifying mildly inequity averse subjects as money-maximizers changes what predictions follow. Fehr and Schmidt must independently justify the 40-60 distribution as one that yields valid predictions.

Second, Fehr and Schmidt (2010, pp 103-104) argue that adherence to the original distribution is not necessary, as their three papers on contract games are not intended as tests of inequity aversion but as rough interpretative exercises—mere illustrations of how fairness models can make sense of contract game behavior. To apply the model as an interpretative exercise takes for granted the very thing that is in question: whether the model deserves explanatory consideration given its severe predictive shortcomings. A few questions follow. Would it not be more appropriate if their papers *were* intended as tests of the model? After all, the data being roughly interpreted should be predictable if inequity aversion were empirically successful—as Fehr and Schmidt must surely hope to demonstrate. Moreover, if their theory and its proper parametric distributions are too inelegant to be useful for this predictive task, should that not count against the usefulness of the theory? The four-type distribution should not be so tedious that the theory's chief advocates and originators cannot use it for prediction.

Third, Fehr and Schmidt (2010) note that their *QJE* distribution is not a unique one, so they do not disregard their ultimatum game findings in changing their distribution. Rather, the distribution they originally declare is merely *suggested* as one among many that are potentially consistent with the ultimatum game data. They acknowledge that "there are many degrees of freedom in choosing a distribution consistent with the ultimatum game" (p 104). This statement serves as explicit confirmation of Binmore and Shaked's (2010) suggestion that their parameters are under-identified. Perhaps the 40-60 distribution is compatible with Fehr and Schmidt's empirical findings, but there are serious problems with the flexibility they exploit in applying their model. In order to confirm or disconfirm a theory, one must be able to derive testable predictions from it, but if the parameters can vary considerably, so too can the consequent predictions. The model is therefore harder to falsify and easier to confirm, as it is consistent with a wide range of observations. Until the theory's proponents are able to pin down parameters, perhaps using additional auxiliary assumptions, the model should not be treated as a viable (or testable) predictive tool, let alone an explanatory one. Even if some are motivated by sheer faith that their work on the model will be vindicated, scholars should adopt a critical lens when assessing Fehr and Schmidt's work. Because Fehr and Schmidt enjoy substantial freedom in choosing parametric distributions, it is plausible that the distributions they have chosen yield more favorable predictions than do alternative distributions that are consistent with the ultimatum game. Fehr and Schmidt (1999) offer few reasons for preferring their distribution other than that it is consistent with the ultimatum game data—like many other distributions seem to be. Since Fehr and Schmidt insufficiently support the distributions they choose, there is no reason to think that the predictions issued from these distributions constitute effective tests of the model's credibility.

A question then arises: what is a more appropriate method of calibrating a model? The inequity aversion model makes a particular assumption about what motivates people. However, these motivations are apt to vary in different contexts. As such, it is methodologically suspect that Fehr and Schmidt derive parametric distributions that fit ultimatum game data and use the same distribution in many other games. Moreover, the extent to which ultimatum game behavior is motivated by inequity aversion is a *psychological* consideration, unknowable through the lens of behavior without surveying social expectations and attitudes. As such, if the explanatory or predictive power of the fairness assumption is to be tested, Fehr and Schmidt ought to find a way to measure degrees of inequity aversion in each experimental context. Contrast this approach with that of social norms theory. Adherents to this paradigm survey experimental subjects to assess empirical expectations, normative expectations, and social expectations, which allow identification of the relevant norms (Bicchieri, 2006). Context- and subject-specific attitudes must be measured for a parametric distribution to be properly set and for the model's assumptions to be crucially tested.

Setting parameters in good faith requires consistency and transparency. Parametric distributions should be set using data and data alone—not set according to the subjective preferences of the authors. Fehr and Schmidt's convenient use of a more elegant or simple distribution trades off with methodological integrity. To the extent that data alone does not dictate the distribution in the same way each time, the model's predictions are *indefinite*. One way that Fehr and Schmidt may crucially test their theory is by sticking to (and defending) their assumption that ultimatum game data reflects a global population distribution of inequity averse attitudes. If predictions then fail, then either this assumption from which parameters are derived must be revised or the utility function itself must be. Alternatively, a different calibration method may be adopted like the one used by the social norms model: attitudinal surveys. Moreover, Fehr and Schmidt are not transparent about how and when they vary their distributions. As noted earlier, they do not show how their joint distribution is derived, nor do they fully explain their reasons for deviating from the ultimatum game distribution until their 2010 response to Binmore and Shaked. For these reasons, I find that Fehr and Schmidt's approach to parameter-setting is scientifically substandard.

## Conclusion

I have discussed three kinds of criticisms that Binmore and Shaked (2010) have leveled against Fehr and Schmidt: favorable selection, predictive inadequacy, and use of floating parameters. Each has weighty implications on how the inequity aversion model should be evaluated. Fehr and Schmidt favorably select events to predict, what to treat as the model's prediction, and what rival theories to compare the model's performance to. Insofar as they do so, the empirical adequacy of the model should remain in doubt, as the model's handle on events that are even remotely challenging to predict has not been proven. Even for the mild predictive challenges that Fehr and Schmidt *do* undertake, their model performs unimpressively. In some cases, its predictions are quite inaccurate. In others, the model performs no better than its alternatives—namely, variants of the money-maximizing model. Finally, Fehr and Schmidt's use of floating parameters raises doubts about the definiteness of their model's forecasts. It is difficult to falsify a model whose predictions are a moving target. Moreover, because Fehr and Schmidt are those who set their parameters' distributions, it is possible that they adjust their distributions in order to generate more favorable predictions. As Binmore and Shaked (2010) note, such a process would entail that Fehr and Schmidt are engaged in a *fitting* exercise as opposed to a *predicting* exercise (p 92). Before they can meaningfully test their theory, Fehr and Schmidt must offer fixed and detailed assumptions from which their parameters' distributions can be precisely determined. In reviewing the responses given by Fehr and Schmidt, I find no satisfactory justifications for these serious methodological and empirical flaws, and, unless these issues have been remedied in their subsequent work, there seems little reason to assign credibility to their model.

## References


Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge UP, Cambridge.

Binmore, K., & Shaked, A. (2010). "Experimental Economics: Where Next?". *Journal of Economic Behavior and Organization* 73, no. 1, 87-100.

Camerer, C., & Thaler, R. (1995). "Ultimatums, Dictators, and Manners". *Journal of Economic Perspectives* 9, 209-219.

Fehr, E., & Gächter, S. (2000). "Cooperation and punishment in public goods experiments". *American Economic Review* 90, 980–994.

Fehr, E., Klein, A., & Schmidt, K. M. (2007). "Fairness and contract design". *Econometrica* 114, 121–154.

Fehr, E., Kremhelmer, S., & Schmidt, K. M. (2005). "Fairness and optimal allocation of property rights". Discussion Paper 5369, CEPR, London.

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

Fehr, E., & Schmidt, K. M. (2003). "Theories of Fairness and Reciprocity: Evidence and Economic Applications". In Dewatripont, S., & Hansen, L. (Eds.), *Advances in Economic Theory: Eighth World Congress* (Volume I), 208–257. Cambridge UP, Cambridge.

Fehr, E., & Schmidt, K. M. (2004). "Fairness and incentives in a multi-task principal-agent model". *Scandinavian Journal of Economics* 106, 453–474.

Fehr, E., Schmidt, K. M. (2005). "The rhetoric of inequity aversion—a reply". <http://www.najecon.org/naj/cache/666156000000000616.pdf>.

Gode, D., & Sunder, S. (1995). "Zero-intelligence traders: Market as a partial substitute for individual rationality". *Journal of Political Economy* 101, 119-137.

Guith, W., Marchand, N., & Rulliere, J. (1997). "On the Reliability o Reciprocal Fairness-An Experimental Study". Discussion Paper, Humboldt University Berlin.

Güth, W., Schmittberger, R., & Tietz, R. (1990). "Ultimatum Bargaining Behavior-A Survey and Comparison of Experimental Results". *Journal of Economic Psychology* 11, 417-449.

Kachelmeier, S. J., & Shehata, M. (1992). "Culture and Competition: A Laboratory Market Comparison between China and the West". *Journal of Economic Organization and Behavior* 19, 145-168.

Kreps, D., Milgrom, P., Roberts, J., & Wilson, R. (1982). "Rational cooperation in the finitely repeated Prisoners' Dilemma". *Journal of Economic Theory* 27, 245–252.

Roth, A. E. (1995). "Bargaining Experiments". in Kagel, J., & Roth, A. (Eds.), *Handbook of Experimental Economics*, Princeton UP, Princeton.

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.

Shaked, A. (2005). "The Rhetoric of Inequity Aversion". UCLA Department of Economics. <https://EconPapers.repec.org/RePEc:cla:levrem:666156000000000570>.

Smith, V. L., & Williams, A. W. (1990). "The Boundaries of Competitive Price Theory: Convergence Expectations and Transaction Costs". in Green, L., & Kagel, J. H. (Eds.), *Advances in Behavioural Economics* (Volume II), Ablex Publishing Corporation, Norwood.

Wilkinson, N. (2008). *An Introduction to Behavioral Economics*. Palgrave Macmillan, Basingstoke, UK.