Behavioral Microeconomics

Ernst Fehr

Report by Jan-Tjeerd Boom, University of Groningen

1. Introduction

Mainstream economics views human beings as being exclusively self-interested. Although many people, including economists, have not felt at ease with this assumption, it was used because it predicted human behavior reasonably. Only lately have economist begun to seriously analyze peoples real behavior in economic situations.

Much of the analysis is done through experiments in which subjects are faced with economic situations. A number of experiments have confirmed the theory of rational, self-interested human behavior. However, in many cases, different behavior was found. In general, people are much nicer and much more cooperative than predicted by the self-interested model. On the other hand are people nastier and even brutal in response to hostile actions than the theory predicts.

In effect, much of human behavior seems to be reciprocal. This means that people are nice to those that are nice to them, while they are willing to punish, even at their own expense, those who do not cooperate.

In his lecture, Fehr gave an overview of some experiments and there outcomes. He first discussed several game theoretic models on which some experiments are based. From this, he derived some basic determinants of human behavior in economic situations and developed a theory that could explain this behavior. Finally, in his last two lectures, he discussed the implications of the design of contracts on behavior of the subjects and tried to explain the evidence with his theory of inequity aversion.

In this report, first a short introduction to laboratory experiments is given. After this, in section 3, an overview of some experiments is given and the theory of inequity aversion is discussed. In section 4 contracts are discussed, and finally in section 5 some conclusions are given.

2. Laboratory Experiments

Fehr defines laboratory economic experiments as follows:

> In laboratory economic experiments, real subjects make decisions in a controlled environment. They are paid real money according to the implemented payoff functions and according to their decisions and – in an interactive situation – according to the decisions of the other subjects.

Doing such experiments has several advantages above observing economic behavior in 'the field'. First, the endogenous and exogenous variables of the economic model can often be measured unambiguously. One reason for this is that there are much fewer unobservable variables. Furthermore, there are no causality problems because 'ceteris paribus' changes can be implemented.
Second, the experimenter often knows the theoretical equilibrium exactly. Therefore, equilibrium and disequilibrium actions can be explicitly observed. This also means that the adjustment process can be explicitly observed and examined.
Third, the experimenter controls the information conditions and the exogenous stochastic processes. This factor is especially important for all models of asymmetric information.
Fourth, the experiment is replicable. Anyone who questions the evidence can replicate the experiment.
Experimental economics is however not without flaws. One criticism is that the setting of the experiment often is very different from the every day experience of the participants. They might therefore react very differently during the experiment than they would in a real economic situation.
A second problem can be that the participants try to react in a way preferred by the experimenter. Even if the experimenter does not explicitly give information on how he expects the results will, or should, be, participants can sometimes find out about this anyway.
Thirdly, the stakes in the experiments are often very low. Most experiments are conducted in industrialized countries where wages are high. The rewards of the experiments are often not more than $10 to $25. It is argued that other incentives than selfish ones may prevail at such low payoffs. Fehr has taken account of this

criticism in many of his experiments by conducting them both in industrialized countries and in developing countries with the same monetary reward. In developing countries, this often means that the stakes are the equivalent of one or more months' wages.

3. Fairness, Competition and Cooperation

The model tested is the standard rational actor model (SRAM), which assumes that economic actors have completely selfish preferences, an unbounded ability to find the best alternative and posses sequential rationality. As the name already says, this is the standard model used in many parts of economics and it can be found in many standard textbooks (see Varian (1992) and Mas-Colell et al. (1995)).
Fehr and Schmidt (1999) review a number of experiments. Some of the models tested confirm the SRAM, while others do not. To explain their empirical findings, they present a new theory on the behavior of economic actors (see also Rabin (1993)). Below I will discuss the different models shortly.

3.1 The Tested Models

Ultimatum Game. The ultimatum game is a two stage game with two players. In the game, one player gives a proposal about the division of a surplus of fixed size. The second player can either accept or reject the proposal. In stage 1, player 1 proposes a share $s \in [0,1]$ to player 2. In the second stage, player 2 either accepts or rejects player 1's offer. The monetary payoffs for both players are (0,0) in the case player 2 rejects the offer and (1-*s*,*s*) if player 2 accepts the offer. The SRAM predicts that player 2 always accepts any $s>0$.
Fehr and Schmidt (1999) find that virtually no offers were made above $s=0.5$ and that in 70% of the cases, offers lay between 0.4 and 0.5 and offers below $s=0.2$ were rarely observed. In stark contrast to the predictions of the SRAM, responders frequently rejected very uneven proposals and the probability of rejection decreases with s. Fehr and Schmidt also found that proposers adjust their proposal to the toughness of the responder, i.e., if the responder is perceived as tough, the proposer offers him a larger share of the surplus. Roth et al. (1991) also found this result.

Market Game with Proposer Competition. In this two stage game, there are *n* players, of which *n*-1 are proposers and 1 player is the responder. In stage 1 the *n*-1 proposers simultaneously propose a share $s_i \in [0,1]$. In stage two, the responder either accepts or rejects the highest s denoted by s^h. The payoffs in both cases are as follows. In the case of rejection by the responder, the payoff is zero for everybody. If the responder accepts s^h, the responder's payoff is s^h, the payoff of the highest proposer is $1-s^h$ and is zero for all the other responders. The equilibrium outcome according to the SRAM is that all proposals are at $s=1$.
In this model, the SRAM is supported. In general, responders accept all $s^h>0.5$ and after a few periods there are many proposals at $s=1$. Furthermore, convergence to the equilibrium is quick.

Market Game with Responder Competition. The market game with responder competition is a three stage game with n players, of which *n*-1 are responders and 1 player is the proposer. In the first stage, player 1 makes a proposal $s \in [0,1]$. In the second game, the *n*-1 responders all individually accept or reject the proposal. If more than one responder accepts the proposal, a third stage is needed, in which nature selects one of the accepting players at random. The monetary payoffs are zero if no responder accepts the proposal. If at least one responder accepts the offer, the payoff of the proposer is 1-*s* and *s* for the selected responder. All other responders receive zero. The SRAM predicts that responders accept any offer $s \geq 0$ and therefore, the proposer offers $s=0$.
Although the predictions of the SRAM are not exactly matched, Güth et al. (1997) find results that come close. Specifically, Güth et al. (1997) find that after four periods, the average acceptance threshold is well below 5% of the surplus. Furthermore, after four periods all offers are below 25% of the surplus and the average offer is about 15% of the available surplus. An interesting finding is that one selfish responder triggers the equilibrium as predicted by the SRAM.

Public Good Game. In the public good game, n players decide simultaneously about their contributions $g_i \in [0, y]$ to the public good, where y is the amount of money available to every player. The monetary payoffs are defined as $x_i = y - g_i + a\sum g_j$, where $a<1<an$. This game is like the classic Prisoners' Dilemma. It does not pay for the individual to contribute to the provision of the public good. However, the

collective payoff would be maximized if all players spend all their money on the public good. The SRAM predicts that all players will free-ride and not donate to the public good.

This is also the result Fehr and Schmidt (1999) arrive at. During the initial periods, average contribution is between 40% and 60%. In the final periods however, the vast majority of subjects free-rides completely.

Public Good Game with Punishment. This game is two stage game. The first stage is equal to the normal public good game. In the second stage, players decide simultaneously whether to punish other players after they observed the contributions $g=(g_1,\ldots,g_n)$ of all players to the public good. The punishment player i inflicts on player j is denoted by p_{ij}. The cost of punishment is $0<cp_{ij}<p_{ij}$. The monetary payoffs are $x_i(g,p)=y-g_i+a\sum g_j-\sum p_{ji}-c\sum p_{ij}$. Note that punishing other players is not individually rational in this game. Hence, the SRAM predicts that players will free-ride and no player will punish.

Fehr and Gächter (1997) find that punishment occurs frequently. In most cases, the contributors to the public good punish the free-riders and the less a player contributes to the public good, the more he is punished. A very surprising finding is that 75% of the subjects contributes the maximum contribution y in the final period.

Gift Exchange Game. The gift exchange game is a two player game with a principal and an agent. The principal offers the agent a certain wage after which the agent either accepts or rejects the offer. If the agent accepts the offer, he receives the wage and exerts a certain effort, e. The payoff of the principal is given by $p=ge-w$, where w is the wage. The payoff of the agent is given by $u=w-c(e)$, where $c(e)$ is the cost of the effort level chosen by the agent. The SRAM predicts that the agent will exert the minimum effort.

In his lecture, Fehr showed that in the gift exchange game, 30% of the subjects act purely selfishly. However, many of the subjects do not behave purely selfishly. In many cases, the effort level of the agents increased in the wage level offered by the principal. This is an example of positive reciprocity, which in general means that people react positively to a positive deed of their opponent.

3.2 Theoretic Considerations

The experiments mentioned above and many others conducted in the past ten years suggest that there exists something as a 'Homo Reciprocans', i.e., cooperative behavior exists. This is not only true when the subjects know each other, but also when the subjects are total strangers. Furthermore, it is also confirmed under experimenter-subject anonymity. This is important with regard to the criticism given above that subjects unconsciously try to live up to the experimenter's expectations. As already mentioned, it is also confirmed when the stakes are rather high and under one-shot repetitions. Still, 'Homo Economicus' also exists. Many experiments suggest that a large fraction of subjects behave purely selfishly.

The interaction between the two different types is of some interest. Positive reciprocity, as in the gift exchange game, induces selfish types to be nice. Negative reciprocity, e.g. punishment for unwanted behavior, deters selfish types to behave opportunistically. This is shown in the public good game with punishment where cooperative subjects through punishment could induce selfish types to contribute to the public good. Finally, selfish types induce homo reciprocans to behave selfishly. In the public good game, one selfish type can induce the cooperative players to defect too.

These findings raise several questions. First of all, do we have to give up the selfishness assumption or the rationality assumption or both? If we give up the selfishness assumption, which motivational forces do we take into account? Several forms of bounded rationality can replace rationality, but which of these do we choose? And last but not least, how do 'non-selfish' motivations interact with bounded rationality?

Fehr suggests the following approach. He is reluctant to give up the rationality assumption, because this seems to explain much behavior in the final period of relatively simple games. At the same time he suggest the idea of inequity, or inequality aversion. Loewenstein et al. (1989) show that at least a fraction of the subjects dislike inequity relative to relevant comparison persons. The aversion against disadvantageous inequity is thereby much stronger that that against advantageous inequity.

Often altruism is given as one explanation for cooperative behavior. According to Fehr, altruism is in the domain of advantageous inequality. On the other hand, relative deprivation, or envy, in the domain of disadvantageous inequality.

4. Contracts

Another conclusion Fehr draws from the experiments mentioned above is that the economic environment, e.g. institutions and contracts, are decisive for what behavior subjects show. In this section, the effect of different contracts on the behavior of subjects will be discussed.

Fehr et al. (1998) and Fehr and Falk (1999) conducted an experiment in which they mimicked a labor market. The labor market was characterized by a large excess supply of labor. In thia case, the SRAM predicts a competitive equilibrium with low wages. Fehr et al. (1998) and Fehr and Falk (1999) found that the outcome depends on the contract offered to the employees. In the presence of complete labor contracts, i.e. contracts in which the effort is exogenously enforced, wages come close to the competitive equilibrium. However, with incomplete labor contracts, i.e. workers can reciprocate, wages stabilize far above the competitive equilibrium. Furthermore, with incomplete labor contracts, competition has no impact on wage formation.

This experiment shows that even with large unemployment, wages do not fall. This is the often documented case of downward wage rigidity. According to Bewley (1995), the main causes of downward wage rigidity have to do with employers' beliefs that other motivators than financial incentives are necessary, which are best thought of as having to do with generosity. Bewley also finds that managers claim that workers have so many opportunities to take advantage of employers that it is not wise to depend on coercion and financial incentives alone as motivators.

The findings made by Bewley (1995) show that most labor relations are based on incomplete contracts. In these contracts, compensation cannot be completely conditioned on workers' efforts and workers have discretion with regard to their effort choice. The result is a motivation problem: workers have an incentive to provide the lowest possible effort level that is compatible with firms' enforcement technology. The main questions can now be raised. Firstly, to what extend does

reciprocity help in overcoming the incentive problem? Secondly, to what extent do explicit economic incentives solve the incentive problem?

These questions are dealt with by Fehr et al. (1996) and Fehr and Gächter (2000). In the experiments, there are two kinds of players; 'firms' and 'workers'. In the first stage, firms make a contract offer existing of *w*, the wage level and $\hat{e}$, the expected level of effort. In the second stage, workers either accept or reject the offer. Upon acceptance, the workers choose their actual effort *e*, which can differ from $\hat{e}$. Hence, the contract (w, $\hat{e}$) is incomplete because *w* is not conditioned on *e*. Therefore, this experiment is called the trust treatment (TT). The monetary payoffs for the firms and workers are respectively:

$$p = 100e - w \qquad (1)$$

$$u = w - c(e) \qquad (2)$$

where $w \in [0,100]$. The prediction is that selfish players will choose the minimum effort level available, which is $e=e^{min}=0.1$, the corresponding equilibrium wage is $w=1$. This is also what the SRAM predicts.

The result of the experiments by Fehr et al. (1996) and Fehr and Gächter (2000) however differed from this prediction. In general much higher wages were offered and workers responded to this by exerting higher effort levels than the absolute minimum. Moreover, the higher the wage offered, the higher the effort of the workers. Hence, firms appeal to workers' reciprocity and many workers respond to this.

Does this change when economic incentives are used instead of appealing to reciprocity? This question was dealt with in the same papers as mentioned above. Now firms could make a contract offer existing of w, $\hat{e}$ and *f*. Here *f* is a fine, constrained by $0 \leq f \leq f^{\max} = 13$, collected by the firm, that the worker has to pay in case of verifiable shirking ($e<\hat{e}$). The workers, upon acceptance of the offer, choose the actual effort level e, which can diverge from $\hat{e}$. In the experiment, a random mechanism determined with probability 1/3 whether $e<\hat{e}$ was verifiable. This incentive contract is more complete than the one discussed above because f is conditioned on e. Therefore, this experiment is called the incentive treatment (IT). The SRAM predicts that a (selfish) worker will perform $\hat{e}$ if $(1/3)f \geq c(\hat{e})$. The maximum fine, f^{max}, determines the maximum enforceable $\hat{e}$.

The question now is, how the reciprocity and the economic incentives interact. As mentioned above, even without incentive contracts, reciprocity gives rise to $e>e^{min}$. For the incentive contract, the SRAM predicts that the real effort level will be equal to the maximum enforceable effort level. There are now two hypotheses. The first one, which is the standard hypothesis, is that financial incentives and reciprocity-driven voluntary co-operation are additive. The second one is that financial incentives crowd out voluntary co-operation.

Fehr and Gächter (2000) give several reasons why incentive contracts might crowd out voluntary cooperation. First of all, reciprocity driven voluntary co-operation may be based on the trust expressed by the principal. The threat to fine the agent is then incompatible with trust. Secondly, the threat to fine the agent may cause a hostile atmosphere. The threat to fine may also destroy the appeal to workers' reciprocity. Finally, principals may now use the stick instead of the carrot, i.e. there are not given any generous offers.

The experiments with the incentive treatment (IT) and trust treatment (TT) gave some interesting results. In the IT, principals rely less on the 'carrot' and more on the 'stick'. On average, wages, rents and desired effort levels are higher in the TT. In the IT, average fines were almost maximal with 69% of all IT contracts having maximal fines. Furthermore, average desired effort levels in the IT are not incentive compatible. In 59% of all IT contracts the demanded effort level is not incentive compatible. A second result is that on average, effort levels in the IT are lower than in the TT. Even in the incentive compatible IT contracts, a large fraction of the agents shirked. Furthermore, voluntary excess effort vanishes completely in incentive compatible IT contracts and in the majority of non-incentive compatible IT contracts, the agents chose the minimum effort level, while in the majority of TT contracts, agents chose non-minimal effort levels. A third result is that IT contracts are less efficient than TT contracts, while principals' profits are highest under incentive compatible IT contracts. The fourth result is that on average, principals in the IT did not offer lower rents at given desired effort levels and in both the IT and TT, principals increased the offered rent if they demanded higher effort levels.

The overall results are that incentive contracts cause a substantial crowding out of voluntary cooperation. This induces firms to make less generous offers and to rely almost exclusively on the ex-ante threat of punishment. Finally, overall efficiency in

the trust treatment is higher, but firm profits are lower, i.e., the introduction of the incentive caused a redistribution in favor of firms and lowered efficiency.

In another paper, Fehr et al. (2000) add another possible contract form; the bunus contract. In the bonus contract, the principal makes a contract offer $(w, \hat{e}, b^*)$. Now there is no fine, but firms can announce to pay a bonus, b^*, in addition to the base wage w, after the firms have observed the actual effort. Note that the bonus announcement b^* is not binding.

So now we have the trust contract, the incentive contract and the bonus contract. What Fehr et al. (2000) wanted to analyze was whether reciprocity affects contractual choices and hence incentive provisions and whether reciprocity gives rise to contractual incompleteness. The experiment conducted consisted of a series of one-shot interactions between principals and agent. The ranking in terms of contractual completeness is as follows. The incentive contract is the most complete, followed by the trust contract, the least complete is the bonus contract. Since the trust contract and the incentive contract were already discussed above, we will concentrate on the bonus contract.

The predictions of the SRAM for the bonus contract are that the actual paid bonus is zero, the actual effort level is e^{min}, only minimal wages are offered by firms in an implicit contract and that the payoffs are the same as in the trust contract. Finally, it is expected that principals prefer the incentive contract over the bonus contract.

The results from the experiments do however show a quite different pattern. First, principals strongly prefer the bonus contract to the incentive contract and the incentive contract over the trust contract. Furthermore, effort is much higher under the bonus contract and somewhat higher under the incentive contract compared to the trust contract. Principals' profits are also much higher under the bonus contract than under the incentive contract. Many principals honour high performance by bonus payments that are in fact conditioned on the effort level. Hence, the opportunity to promise and pay bonuses has a strong incentive effect.

As mentioned above, Fehr together with others has developed a theory on inequity aversion. The question is, whether this theory can explain the major behavioral patterns. In their paper, Fehr et al. (2000) show that they can for the most part. I will not give the technical details of this analysis, but only the intuition behind the explanations.

Inequity averse principals can and do condition the bonus payment on the observed effort level of workers. This provides strong pecuniary incentives for the selfish agents to perform under the implicit contract. Inequity averse agents shirk, because they are afraid of being cheated by the selfish principals. Since the average effort is higher in the implicit contract, inequity averse principals prefer the implicit contract. Therefore, selfish principals also prefer the implicit contract, because they enjoy the benefits (high effort) without the costs (zero actual bonus).

The conclusions of this section are that experiments indicate a strong preference for the implicit, least complete, contract. Furthermore, the theory of fairness developed by Fehr and Schmidt (1999) is consistent with the major facts and offers the following explanations. Fair principals, because of their ability to keep 'non-binding' promises, use implicit contracts to provide strong incentives for selfish agents. Selfish principals prefer implicit contracts because the presence of fair principals allows them to cheat the agents. Finally, fair agents shirk because they dislike being cheated.

5. Conclusions

The analysis discussed above suggests that 'Homo Economicus', defined as a rational and selfish actor, exists. However, besides homo economicus, there also exists 'Homo Reciprocans'. Moreover, 'Homo Reciprocans' seems to be in the majority. The existence of 'cooperative man' is supported by many experiments. It is shown that subjects behave cooperatively even to total strangers, when the stakes are high and under one-shot repetitions. Furthermore, it is also confirmed under experimenter-subject anonymity. However, it depends on institutional factors, e.g. contract design, which behavior prevails.

To model such behavior, Fehr argues that the concept of rationality should not be given up. However, the selfishness assumption should be replaced, at least for a part of the population by a inequity aversion assumption. Inequity aversion means that people receive disutility from differences in well being between subjects.

In the part on contracts it was shown that incentive contracts crowd out voluntary cooperation. This induces firms to make less generous offers and to rely on the ex-ante threat of punishment. Furthermore, the incentive is less efficient, but results in a higher profit for the firms. Compared with both incentive and trust contracts, bonus

contracts perform best, even though they are the contractual completeness of this contract is lowest of all three. Principals strongly prefer the bonus contract to the incentive and trust contract. The reason is that inequity averse principals can through bonus contracts condition the bonus payment on the observed effort level of workers. This provides an incentive for workers to exert a high effort level. In turn, selfish principals also want to use bonus contracts because it gives them a high effort and they can defect on the payment of the bonus. Because inequity averse workers are afraid of this kind behavior, they shirk.

This also shows that the theory of inequity averseness can predict the behavior of a large part of the population. Although it does not predict behavior right in every instance, it is an important improvement on the traditional standard rational actor model.

Literature

Bewley, T. (1995), A Depressed Labor Market as Explained by Participants, *American Economic Review, Papers and Proceedings 85*, 250-254.

Fehr, E. and Falk, A. (1999), Wage Rigidity in a Competitive Incomplete Contract Market, *Journal of Political Economy 107-8*, 106-134.

Fehr, E. and Gächter, S. (2000), Do Incentive Contracts Crowd Out Voluntary Cooperation?, Working Paper 34, Institute for Empirical Economic Research, University of Zürich.

Fehr, E., Gächter, S. and Kirchsteiger, G. (1996), Reciprocity as Contract Enforcement Device, *Econometrica 4*, 833-868

Fehr, Ernst, Klein, Alexander and Schmidt, Klaus (2000), Endogenous Incomplete Contracts, mimeo, University of Munich.

Fehr, Ernst and Schmidt, Klaus (1999), A Theory of Fairness, Competition and Cooperation, *Quarterly Journal of Economics 114* (3), 817-868

Güth, W., Marchand, N. and Rulliere, J.-L. (1997), On the Reliability of Reciprocal Fairness – An Experimental Study, Working Paper, Humboldt University, Berlin.

Loewenstein, G. Thompson, L. and Bazerman, M. (1989), Social Utility and Decision Making in Intertemporal Contexts, *Journal of Personality and Social Psychology 3*, 426-441.

Mas-Colell, Andreu, Whinston, Micheal D. and Green, Jerry R. (1995), *Microeconomic Theory*, Oxford University Press, Oxford

Rabin, Matthew (1993), Incorporating Fairness into Game Theory and Economics, *American Economic Review 83* (5), 1281-1302.

Roth, A., Prasnikar, V., Okunu-Fujiwara, M. and Zamir, S. (1991), Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An Experimental Study, *American Economic Review 81*, 1068-1095

Varian, Hal R. (1992), *Microeconomic Analysis*, Third Edition, W.W. Norton, New York.

--