Erlei, Mathias

Working Paper

Heterogeneous Social Preferences

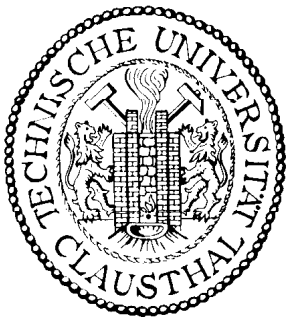
TUC Working Papers in Economics, No. 01

Provided in Cooperation with:
Clausthal University of Technology, Department of Economics

Suggested Citation: Erlei, Mathias (2003) : Heterogeneous Social Preferences, TUC Working Papers in Economics, No. 01, Technische Universität Clausthal, Abteilung für Volkswirtschaftslehre, Clausthal-Zellerfeld, http://dx.doi.org/10.21268/20140612-234057

This Version is available at:
http://hdl.handle.net/10419/107442

Clausthal University of Technology



Heterogeneous Social Preferences

Mathias Erlei [*, #]

This version: June 2004

Abstract: Recent research has shown the usefulness of social preferences for explaining behavior in laboratory experiments. This paper demonstrates that models of social preferences are particularly powerful in explaining behavior if they are embedded in a setting of heterogeneous actors with heterogeneous (social) preferences. For this purpose a simple model is introduced that combines the basic ideas of inequity aversion, social welfare preferences, reciprocity and heterogeneity. This model is applied to 43 games and it can be shown that its predictive accuracy is clearly higher than that of the isolated approaches. Furthermore, it can explain most of the "anomalies" (the "contradictions") that are discussed in Goeree and Holt (2001).

JEL-Classification: C72, C92, D63

---

[*] Institute of Business Administration and Economics, Clausthal University of Technology, Julius-Albert-Str. 2, 38678 Clausthal-Zellerfeld, Germany, m.erlei@tu-clausthal.de.

[#] For helpful comments I wish to thank Heike Schenk-Mathes, J. Philipp Siemer and Jens-Peter Springmann.

1. Introduction

When we meet people for the first time in our life we often ask ourselves what *kind* of person he or she is. And it is not just a question of curiosity that we like to know what type of person we are acquainted with. Most often this is *the* central question for our decision to have further contact with that person. Such a way of thinking suggests that people *are* indeed very different from each other and that the *type* of the person we are occupied with is of uttermost importance. However, economic theory, by and large, ignores differences between people and usually assumes homogeneous preferences. Maybe economic theory misses too much of human behavior by doing so. The main purpose of this paper is to demonstrate the usefulness of explicitly modeling heterogeneity in preferences for explaining the behavior of subjects in 43 laboratory experiments. Next to this we try to corroborate the relevance of social preferences for explaining human behavior.

Experimental evidence makes clear that there are many games in which Nash Equilibrium describes people's behavior quite well. However, there seem to be just as many other games in which laboratory behavior deviates from the predictions of standard game theory by a wide margin. Obviously, there is a need for theoretical innovations which can explain the successes of game theory as well as its failures. No doubt, theory has reacted to experimental evidence. There are several branches of new theoretical approaches that can claim to have at least partial success in introducing superior concepts. Dynamic evolutionary approaches[1] (e.g. replicator dynamics) often but not always converge to Nash Equilibria. Quantal Response Equilibria in general and the Logit Equilibrium in particular (McKelvey and Palfrey 1995, 1998 and Goeree and Holt 2001) have been quite successful in explaining behavioral reactions due to parameter variations in games with identical Nash Equilibria. Finally, there is a third strand of research which was successful in explaining deviations from Nash Equilibrium. These are approaches of "other regarding" or "social" preferences. The social preferences approach can be divided into at least three important substrands: theories of intentional reciprocity (Rabin 1993, Dufwenberg and Kirchsteiger 1998), the inequity aversion approach (Bolton and Ockenfels 2000, Fehr and Schmidt 1999) and recently a theory of Social Welfare Preferences

[1] See Weibull (1995) or Fudenberg and Levine (1998).

(Andreoni and Miller 2002 and Charness and Rabin 2002). In this paper we shall concentrate on the last two approaches.

Bolton and Ockenfels (2000) as well as Fehr and Schmidt (1999) introduce concepts of inequity aversion. It is assumed that there exist people who dislike inequality and who actually sacrifice money to reduce it. Both concepts are particularly successful in describing laboratory behavior when they assume heterogeneous actors. Bolton's and Ockenfels' model is exclusively defined for heterogeneous populations of subjects. Although Fehr's and Schmidt's model can be used for homogeneous populations, all successful applications assume a mixture of inequity averse and strictly egoistic subjects, the latter being individuals with the standard utility functions in game theory. The approaches differ in the concrete definition of inequity aversion and Bolton and Ockenfels allow for more general preference distributions of subjects. However, the general version of their model is somewhat more complicated and this makes it less suitable for direct application. It thus cannot surprise very much that most further applications of the inequity aversion approach use the simpler Fehr and Schmidt variant. In the meantime inequity aversion has been challenged by numerous experiments that have been carried out (e.g. Kagel and Wolfe 2000 and Charness and Rabin 2002).

The most recent alternative to inequity aversion has been presented by Charness and Rabin (2002). They introduce a model of social welfare preferences with and without reciprocity. Social welfare preferences are characterized by individuals who give positive weight to aggregated surplus, i.e. if other people are better off, c.p., utility of individuals increase. The authors carried out 32 experiments and compared the compatibility of several social preference approaches with the experimental data. Their conclusion is that social welfare preferences show the best fit to the data. However, the comparison between social welfare preferences and the inequity aversion model is biased because Charness and Rabin do not take into account that the most fruitful version of the inequity aversion model takes explicitly into account that there are different types of actors, i.e. that there is heterogeneity of preferences. In fact, in Fehr, Kremhelmer and Schmidt (2002) as well as in Fehr, Klein and Schmidt (2001) inequity averse actors are only a minority in the population and the explanatory power of the model stems in particular from the interplay of strictly egoistic and inequity averse subjects. However, Charness and Rabin (2002) only consider the homogeneous population variant of the inequity aversion theory.[2] The same critique also applies to their own model of social wel-

[2] Charness and Rabin (2002) are well aware of this shortcoming as footnote 6 of their paper shows.

fare preferences in which they assume a monomorphic population, again. In this paper we shall try to show that this shortcoming seriously limits the explanatory power of their model. Nevertheless, Charness and Rabin convincingly show that social welfare preferences might help explaining quite a lot of behavior in their 32 games.

Summarizing, Fehr and Schmidt have shown the usefulness of modeling heterogeneous population equilibria with inequity averse and strictly egoistic agents. Charness and Rabin have shown some evidence for social welfare preferences and the relevance of reciprocity. This paper tries to combine these approaches and analyzes whether this increases explanatory power significantly. The focus in this paper is on the application of the basic idea. Therefore, the basic model has to be sufficiently tractable for direct application in a wide variety of games. In fact, we are going to apply the model to 43 different games and show that its predictive accuracy is clearly greater than that of the isolated models.

However, the reader should be well aware that the model presented in this paper is regarded just as one single step to the development of *operational* models for explaining experimental and field evidence. Its main purpose is to demonstrate the importance of heterogeneity of preferences.

In section 2 we shall introduce a very simple 2 players – 3 types model of heterogeneous preferences with explicit modeling of reciprocity. In section 3 this model is applied to all 2-player experiments in Charness and Rabin (2002). Furthermore, its predictive accuracy is compared with that of the inequity aversion and the social welfare preference approach. In section 4 the model is applied to eight games (each game is analyzed for 2 variants with identical Nash Equilibria) from Goeree and Holt (2001). Finally, a summary, some conclusions and some thoughts about future research are given in section 5.

2. Heterogeneous Social Preferences: A Simple Model

In this section I outline a simple 2 players model with linear objective functions. The main purpose of the model is to combine elements of the approaches of Fehr and Schmidt (1999) and of Charness and Rabin (2002) in a rather simple way that allows for direct application to 2 person games. The main ingredients from Charness and Rabin (2002) are the concept of Welfare Preferences and negative reciprocity. The ideas taken from Fehr and Schmidt (1999) are mainly that there are different types of actors, i.e. there is heterogeneity among the players, and the concept of inequity averse players. It is assumed that there are three kinds of players: strictly egoistic actors (SE actors), inequity averse actors (IA actors) and one type of

individuals who has Welfare Preferences (WP actors). It is further assumed that there is incomplete information in the sense that individuals only know their own types but not the types of the actors they are playing with. However, they do know the distribution of types among all individuals in their society so that they have common priors about the distribution of types.

In accordance with Charness and Rabin (2002) we do not explicitly take into account positive reciprocity. We assume that all effects of positive reciprocity are represented by IA and WP preferences. Charness and Rabin report that they find only little evidence for positive reciprocity. Consequently, there is some hope that the neglect of positive reciprocity does not lead to high inaccuracies.

In contrast to positive reciprocity, negative reciprocity is modeled explicitly. If other players "misbehave" parameters for social preferences in IA and WP utility functions are changed so that players become more "envious" and less generous, respectively. The corresponding utility functions are given by equation (1)

$$U_i(\pi_i,\pi_j)=\begin{cases}(1-\sigma_t-\theta_t R)\pi_i+(\sigma_t+\theta_t R)\pi_j, & \pi_i<\pi_j\\ (1-\rho_t-\theta_t R)\pi_i+(\rho_t+\theta_t R)\pi_j, & \pi_i>\pi_j\\ (1-\theta_t R)\pi_i+\theta_t R\pi_j, & \pi_i=\pi_j\end{cases} \quad (1)$$

$i = 1, 2$, $i \neq j$, t = SE, IA, WP. Here, π_i is the monetary payoff of player i. ρ_t represents player i's concern for player j's payoff if player i's payoff is larger than player j's. σ describes the weight that player i puts on player j's payoff if player j gets the higher payoff. Finally, θ is the reciprocity parameter and R is the reciprocity variable which is –1 if the other player has misbehaved and it is zero if no misbehavior has occurred.

Strictly egoistic players (SE) are characterized by $\sigma_{SE} = \rho_{SE} = \theta_{SE} = 0$, i.e. they just care about themselves. Thus, SE players have standard game theoretic utility functions. Inequity averse players' (IA) parameters are characterized by $\sigma_{IA} < 0 < \rho_{IA} < 1$. This means that they put negative weight on the other player's payoff if their own payoff is lower and that they put a positive weight on it if the own payoff is higher than the other player's payoff: IA players dislike inequity. WP players always put a positive weight on the other player's payoff. It is assumed that $0 < \sigma_{WP} \leq \rho_{WP} \leq 1$. Furthermore, it is assumed that $\theta \geq 0$ for all types.

The reciprocity variable R is dependent on "misbehavior" of the other player. Therefore, we have to define misbehavior. Player i regards player j's action as misbehavior if j's action violates i's "norm" *and* if i cannot make sure that his final utility is at least as large as the one he would have got if j had acted according to i's norms. Let us assume that different types of

players have different norms. Furthermore, it is important that each player is willing to behave according to his own norms if he knows that the other actor follows the same norms. Otherwise such norms would be eroded quickly. Consequently, norms have to be best responses in a game in which the player is playing against another player who has identical norms. As this has to be true for both players we can now define norms.

Definition 1: *The norm behavior for players of type t is defined by the (Subgame Perfect) Nash Equilibria of the corresponding complete information game (with IA or WP preferences, respectively, and R = 0) in which two players of the same type play against each other. The corresponding equilibria shall be called "normative reference equilibria".*

This means that players always follow the norms if they *know* that the other player belongs to the same type of actors as they themselves. However, if they do *not* know the other player's type they may deviate from their own norm. It is important to remember that deviating from one's norm is only a necessary (but not a sufficient) condition for "misbehavior". Imagine a situation in which actor j has deviated from i's norm but in which player i can guarantee himself an even higher utility than in the "normative reference equilibrium". In this case he has no reason to be angry and to punish j. In such a case it is rather implausible that deviation from the norm will trigger any negative reciprocity. To exclude such cases from triggering reciprocity we add a maximin condition: Negative reciprocity ($R = -1$) is only triggered if a deviation from the norm occurs *and* the maximin utility for player i in the remaining game is smaller than his utility in the "normative reference equilibrium". Otherwise $R = 0$.

The existence of different types of players with different utility functions means that the games we are going to analyze are games with incomplete information. Therefore, it seems natural to apply standard tools of game theory. Let the "material game" be the standard representation of a game with payoffs only in monetary units. Then the "heterogenous utility game" of the original material game is the extended version of the latter into a game with incomplete information in which three types of players, SE, IA and WP subjects, with known shares in the population exist and payoffs are given in type specific utilities.

Definition 2: *Heterogeneous Social Preferences* (*HSP) Equilibria are given by the (perfect Bayesian) Equilibria of the heterogeneous utility game.*

Note that we have defined IA and WP preferences only for 2-player games yet. Consequently, HSP Equilibria according to Definition 2 are also only defined for 2-player games. Although extensions to n player definitions are not difficult there are several possible ways to extend IA

and WP preferences to more general settings. As this is not necessary for the analysis in the remaining part of this paper, such generalizations remain objectives for future work.

Let us now turn to the 2 players approach, again. In the following sections we shall show that this simple extension of the established models leads to a surprising increase in predictive accuracy.

3. HSP Equilibria in the Charness-Rabin-Games

In this section and in section 4 we will apply the HSP Equilibrium concept to 43 different games. The first 27 games are the 2-players games that are presented in Charness and Rabin (2002). The remaining games are 8 "treasure games" which confirm traditional Nash Equilibria and 8 "contradiction games" that contradict standard Nash predictions taken from Goeree and Holt (2001).

To be able to derive concrete predictions, we have to make some further assumptions concerning the parameters in the IA and WP utility functions. In their recent papers Fehr and Schmidt successfully use $\sigma_{IA} = -2$ and $0 < \rho_{IA} < 1$ to explain laboratory behavior. As we try to stick to their IA model rather closely we adopt σ_{IA} and assume that $\rho_{IA} = 0.75$. To keep the analysis as simple as possible and to give reciprocity a sufficient importance assume that in case of negative reciprocity IA players always behave as if they had less money than the other player, i.e. they put negative weight on the other player's payoff:

$$\theta_{IA} = \theta_{IA}(\pi_i, \pi_j, R) = \begin{cases} \rho_{IA} - \sigma_{IA} & \text{if } R = -1 \text{ and } \pi_i > \pi_j \\ -\sigma_{IA} & \text{if } R = -1 \text{ and } \pi_i = \pi \\ 0 & \text{otherwise} \end{cases}$$

so that in case of $R = -1$ the sum of σ and θ as well as the sum of ρ and θ always equals –2. The corresponding utility function of player i thus is

$$U_i^{IA}(\pi_i, \pi_j, R) = \begin{cases} 3\pi_i - 2\pi_j & \text{if } \pi_i < \pi_j \text{ or } R = -1 \\ 0.25 \cdot \pi_i + 0.75 \cdot \pi_j & \text{if } \pi_i > \pi_j \text{ and } R = 0 \\ \pi_i & \text{if } \pi_i = \pi_j \text{ and } R = 0 \end{cases}.$$

With regard to the WP parameters Charness and Rabin (2002) remain rather vague. Although they do estimate these parameters – in their estimation with the best fit they get $\sigma_{WP} = 0.023$ and $\rho_{WP} = 0.424$ – the results give us only limited guidance for our approach because they assume homogenous actors. However, as there is not a single estimation with $\sigma_{WP,}\ \rho_{WP} > 0.5$

we regard 0.5 as an upper boundary of these parameters. Furthermore, they assume $\sigma_{WP} \leq \rho_{WP}$. Finally, to get a qualitative distinction between ρ_{WP} and σ_{WP} we assume that $\rho_{WP} = 0.5$ and $\sigma_{WP} = 0.3$ which works out quite well in the following applications. Such a value of ρ_{WP} expresses the idea of "Welfare Preferences" most closely as it gives equal weights to both players' payoffs. Furthermore, it seems quite plausible that players give less weight to the other player's payoff if they get less money than the other one. With regard to reciprocity we assume that in case of negative reciprocity WP subjects behave like SE actors, i.e. they are strictly egoistic. This corresponds quite nicely to Charness' and Rabin's idea of 'concern withdrawal': "they withdraw their willingness to sacrifice to allocate the fair share toward somebody who himself is unwilling to sacrifice for the sake of fairness." Consequently,

$$\theta_{WP} = \theta_{WP}(\pi_i, \pi_j, R) = \begin{cases} \sigma_{WP} & \text{if } \pi_i < \pi_j \text{ and } R = -1 \\ \rho_{WP} & \text{if } \pi_i > \pi_j \text{ and } R = -1 \\ 0 & \text{if } \pi_i = \pi_j \text{ or } R = 0 \end{cases} \qquad \text{so that}$$

$$U_i^{WP}(\pi_i, \pi_j, R) = \begin{cases} 0.7 \cdot \pi_i + 0.3 \cdot \pi_j & \text{if } \pi_i < \pi_j \text{ and } R = 0 \\ 0.5 \cdot \pi_i + 0.5 \cdot \pi_j, & \text{if } \pi_i > \pi_j \text{ and } R = 0. \\ \pi_i, & \text{if } \pi_i = \pi_j \text{ or } R = -1 \end{cases}$$

Of course, these parameters are rather crude and subjective first estimates which only serve as a first step in developing a model of heterogeneous social preferences.

Finally, we have to make assumptions about the distribution of types. A rough estimate[3] of the Charness-Rabin games is that strictly egoistic actors amount to approximately 50% of subjects, inequity averse players make up about 15% and the remaining 35% of the players are WP actors.

For the purpose of better intuition about the nature of HSP Equilibria let us first discuss the equilibrium in one of the 27 games, game Barc1 (which is identical to Berk13), more explic-

[3] The "rough estimate" was carried out the following way: First, decisions in the first seven games of Charness and Rabin (2002) – which are pure dictator decisions – were determined and then an estimation of population shares was carried out. Then equilibria of all 27 games were determined assuming this distribution (with generously rounded values of the population shares). Next, another estimation of population shares was carried out. This procedure was repeated twice. Finally, the population shares were, again, generously rounded so that we get prominent numbers as population shares.

itly. The game consists of two stages. In stage 1 player A chooses monetary payoffs of 550 for both players, (550, 550), or he lets player B choose between (400, 400) and (750, 375).[4] If $R = 0$ then the corresponding utilities of the three types of players are given by[5]

	$U^{SE}(R = 0)$		$U^{IA}(R = 0)$		$U^{WP}(R = 0)$	
Out	**550, 550**		**550, 550**		**550, 550**	
Enter	**400, 400**	**750, 375**	**400, 400**	**468.75, –375**	**400, 400**	**562.5, 487.5**
	Left	*Right*	*Left*	*Right*	*Left*	*Right*

Table 1: Utilities in Barc1/ Berk13

First, we have to determine the "normative reference equilibria", i.e.the subgame perfect equilibria in the complete information game where an IA player plays with an IA player (or a WP player playing with a WP player). As can easily be seen, inequity averse B players prefer *Left* and, anticipating this, inequity averse A players thus choose *Out*. The normative reference equilibrium of IA subjects then is (*Out*, *Left*). Analogously, B players with WP prefer *Right* and WP-A players choose *Enter* because 562.5 > 550. The normative reference equilibrium of WP players thus is (*Enter*, *Right*).

The normative reference points determine whether negative reciprocity will be triggered. Because WP players regard *Enter* as part of their norm and as *Out* immediately finishes the game WP subjects cannot show negative reciprocity in this game. However, inequity averse subjects regard *Out* as the normatively adequate behavior. Furthermore, the maximin value of inequity averse B players after *Enter* equals 400 which is strictly less than their utility in the normative reference equilibrium. Consequently, if A chooses *Enter* this will trigger their negative reciprocity reaction. However, in this game utilities of inequity averse players do not change if $R = -1$, as can easily be checked.

Knowing this, we can determine equilibrium behavior of all three types in stage 2. SE and IA actors choose *Left* and WP subjects choose *Right*. Given the distribution of the types this means that the probability that player B plays *Left* is 0.65. Next we can calculate expected utility of player A choosing *Enter*. For SE players EU(*Enter*) = 0.65×400 + 0.35×750 = 522.5

[4] The first number in parentheses corresponds to the monetary payoff of player A, the second number to B's payoff.

[5] One can read the following table in the following way, too: In stage 1 of the game: player A chooses a row (*Out* or *Enter*) and in stage 2 player B chooses a column (*Left* or *Right*).

which is less than 550, his utility when choosing *Out*. Consequently, SE players choose *Out*. As can easily be checked, the same is true for IA and WP players. Consequently, the unique HSP Equilibrium consists of all A players opting for *Out* and 65 percent of B players choosing *Left*. By and large, this equilibrium is confirmed by experimental behavior. In experiment Barc1 (Berk13) 96 percent (86 percent) of A players have chosen *Out* and 93 percent (82 percent) of B players have chosen *Left*.

In the same way all 2 player experiments in Charness and Rabin (2002) have been analyzed. In Table 2 the structure of all 27 experiments is summarized. P(*Enter*) (P(*Left*)) represents the percentage of A players (B players) that have decided to let B choose (that have chosen *Left*). The corresponding equilibria are given in Table 3.

#	Name	Experiment / Game	P(*Enter*)	P(*Left*)
1	Berk29	B chooses (400,400) vs. (750,400)		.31
2	Barc2	B chooses (400,400) vs. (750,375)		.52
3	Berk17	B chooses (400,400) vs. (750,375)		.50
4	Berk23	B chooses (800,200) vs. (0,0)		1.00
5	Barc8	B chooses (300,600) vs. (700,500)		.67
6	Berk15	B chooses (200,700) vs. (600,600)		.27
7	Berk26	B chooses (0,800) vs. (400,400)		.78
8	Barc7	A chooses (750,0) or lets B choose (400,400) vs. (750,400)	.53	.06
9	Barc5	A chooses (550,550) or lets B choose (400,400) vs. (750,400)	.61	.33
10	Berk28	A chooses (100,1000) or lets B choose (75,125) vs. (125,125)	.50	.34
11	Berk32	A chooses (450,900) or lets B choose (200,400) vs. (400,400)	.15	.35
12	Barc3	A chooses (725,0) or lets B choose (400,400) vs. (750,375)	.26	.62
13	Barc4	A chooses (800,0) or lets B choose (400,400) vs. (750,375)	.17	.62
14	Berk21	A chooses (750,0) or lets B choose (400,400) vs. (750,375)	.53	.61
15	Barc6	A chooses (750,100) or lets B choose (300,600) vs. (700,500)	.08	.75
16	Barc9	A chooses (450,0) or lets B choose (350,450) vs. (450,350)	.31	.94
17	Berk25	A chooses (450,0) or lets B choose (350,450) vs. (450,350)	.38	.81
18	Berk19	A chooses (700,200) or lets B choose (200,700) vs. (600,600)	.44	.22
19	Berk14	A chooses (800,0) or lets B choose (0,800) vs. (400,400)	.32	.45
20	**Barc1**	A chooses (550,550) or lets B choose (400,400) vs. (750,375)	.04	.93
21	**Berk13**	A chooses (550,550) or lets B choose (400,400) vs. (750,375)	.14	.82
22	Berk18	A chooses (0,800) or lets B choose (0,800) vs. (400,400)	1.00	.44

23	Barc11	A chooses (375,1000) or lets B choose (400,400) vs. (350,350)	.46	.89
24	Berk22	A chooses (375,1000) or lets B choose (400,400) vs. (250,350)	.61	.97
25	Berk27	A chooses (500,500) or lets B choose (800,200) vs. (0,0)	.59	.91
26	Berk31	A chooses (750,750) or lets B choose (800,200) vs. (0,0)	.27	.88
27	Berk30	A chooses (400,1200) or lets B choose (400,200) vs. (0,0)	.23	.88

Table 2: The Charness-Rabin experiments / games.
P(Left): Percentage that subjects choose "Left"; P(Enter): Percentage that subjects choose "Enter".

#	Name	Player A P(Enter)			Player B P(Left)			P(*Enter*)	P(*Left*)
		SE	IA	WP	SE	IA	WP		
1	Berk29				[0,1]	1	0		[0.15,0.65]
2	Barc2				1	1	0		0.65
3	Berk17				1	1	0		0.65
4	Berk23				1	0	1		0.85
5	Barc8				1	1	0		0.65
6	Berk15				1	0	0		0.50
7	Berk26				1	0	[0,1]		[0.50,0.85]
8	Barc7	0	1	1	[0,1]	1	0	0.50	[0.15,0.65]
9	Barc5	1	0	0	[0,118/140]	1	0	0.50	[0.15,0.57]
		0	0	0	[118/140,1]	1	0	0	[0.57,0.65]
		[0,1]	0	0	118/140	1	0	[0,0.50]	0.5714286
10	Berk28	1	1	0	[0,1]	0	[0,1]	0.65	[0,0.50]
		0	1	0	[0,1]	0	[0,1]	0.15	[0.50,0.85]
		[0,1]	1	0	[0,1]	0	[0,1]	[0.15,0.65]	0.50
11	Berk32	0	1	0	[0,1]	0	[0,1]	0.15	[0,0.85]
12	Barc3	0	1	1	1	1	0	0.5	0.65
13	Barc4	0	1	1	1	1	0	0.50	0.65
14	Berk21	0	1	1	1	1	0	0.50	0.65
15	Barc6	0	0	1	1	1	0	0.35	0.65
16	Barc9	0	1	1	1	1	1	0.50	1
17	Berk25	0	1	1	1	1	1	0.50	1
18	Berk19	0	0	1	1	0	0	0.35	0.50
19	Berk14	0	0	0	1	0	[0,1]	0	[0.50,0.85]
20	**Barc1**	0	0	0	1	1	0	0	0.65
21	**Berk13**	0	0	0	1	1	0	0	0.65
22	Berk18	1	1	1	1	0	[0,1]	1	[0.50,0.85]
23	Barc11	1	1	0	1	1	1	0.65	1
24	Berk22	1	1	0	1	1	1	0.65	1
25	Berk27	1	0	0	1	0	1	0.50	0.85
26	Berk31	0	0	0	1	0	1	0	0.85
27	Berk30	0	1	0	1	0	1	0.15	0.85

Table 3: HSP Equilibria in the Charness-Rabin games

How well do these equilibria explain laboratory behavior? To answer this question we first have decide how to handle multiple equilibria. Table 3 shows that in 12 of 27 games there are, indeed, multiple equilibria. Of course, some of these equilibria explain behavior better than other equilibria of the same game. We analyze three "scenarios: (a) We only take into account the "best" equilibria. By this we mean those equilibria that have the smallest mean absolute error (MAE) over both the *Entry-Out* and the *Left-Right* decision. (b) According to the same standard we take into account only the worst equilibria and (c) we separately analyze those games which have a unique equilibrium.

Figure 1 shows the percentage of subjects playing *Left* or *Enter* (P_Left) and the corresponding best case probabilities of *Left* (cases 1 to 27) and *Enter* (cases 28 to 47) in HSP equilibria (Prog_HSP).

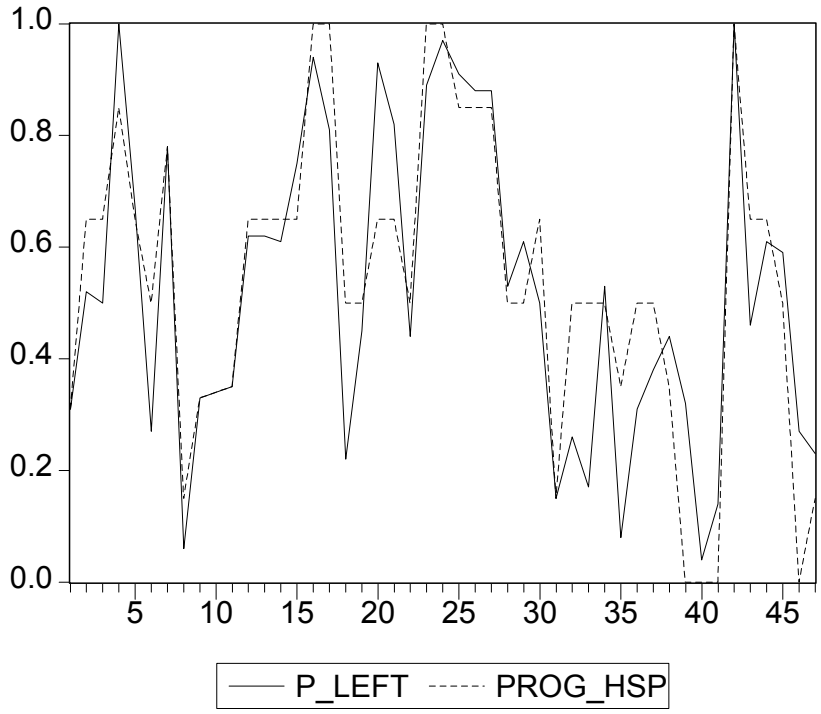


Figure 1: HSP predictions and laboratory behavior

Figure 1 seems to indicate that HSP covers the main qualitative features of the behavior in the Charness-Rabin games. This impression is further strengthened by some statistical measures of predictive accuracy. Now let $R^2 = 1 - \left(\sum (PPL - PL)^2 \Big/ \sum (PL - \mu)^2\right)$ be the "quasi-coefficient of determination" with *PPL* as the predicted probability of choosing *Left* (or *Enter*), *PL* as the actual percentage of *Left* (or *Enter*) and μ as the mean percentage of *Left* (or *Enter*) over all strategies. Then $R^2 = 0.73$. Furthermore, the mean absolute error of the HSP Equilibrium is $MAE = 0.1074$.

If we take the "best" Nash Equilibrium as an estimator of laboratory behavior predictive accuracy is clearly lower (see Figure 2). The quasi-coefficient of determination is even negative ($R^2 = -0.51$) and the mean absolute error is much higher ($MAE = 0.263$). Obviously, Nash Equilibria are a rather bad estimator of actual laboratory behavior.

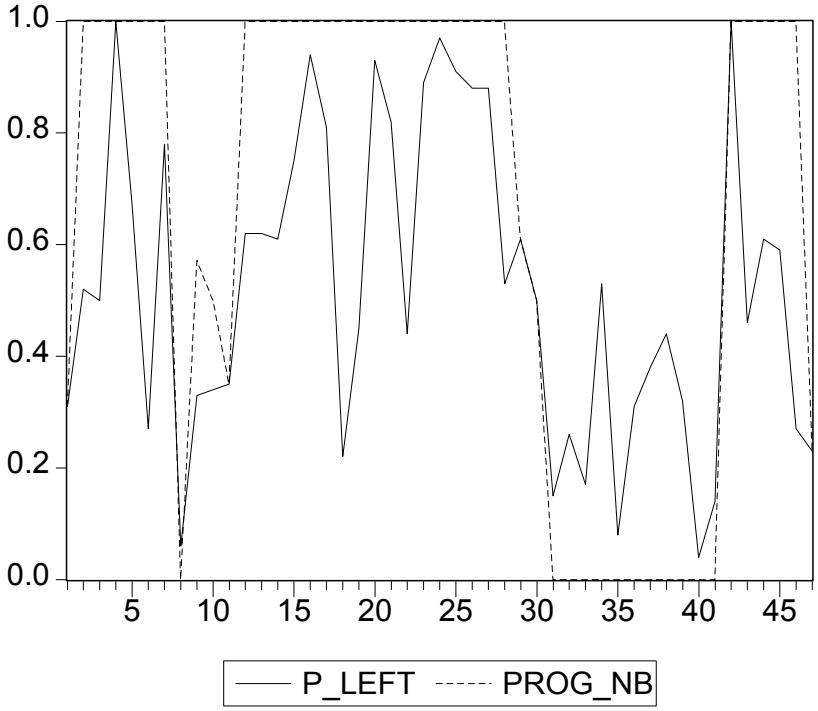


Figure 2: Subgame Perfect Nash Equilibrium prediction and laboratory behavior

The Fehr-Schmidt model of inequity aversion works a little better than Nash Equilibrium. Here we take parameters that the authors use in Fehr, Klein and Schmidt (2001) and Fehr Kremhelmer and Schmidt (2002), i.e. $\sigma_{IA} = -2$, $\rho_{IA} = 0.75$ and assume that 60 percent of subjects are SE actors and that 40 percent of subjects are inequity averse. If, again, we take only the "best" equilibria according to the Fehr et al. model we get $R^2 = 0.033$ and $MAE = 0.22$. The model clearly works worse than the concept of HSP Equilibria.

To test the Charness-Rabin model of a population consisting only of (homogenous) WP individuals we assume that $\sigma = 0.023$, $\rho = 0.424$ and $\theta = 0.111$.[6] In addition, for this test we accept Charness' and Rabin's (2002, p. 840) definition of misbehavior, i.e. entry by A is characterized as misbehavior in games 9, 11, 19, 20, 21, 23, 24, 25, 26, 27 of Table 2. Predictions of behavior are determined by the subgame perfect Nash Equilibrium given the Charness-Rabin

[6] These are the values of Charness' and Rabin's best estimate. See Charness and Rabin (2002), p. 840.

utility function. It shows that the Charness-Rabin model does not work very well. $R^2 = -1.096$ and $MAE = 0.33$ are both worse than the corresponding values of the Fehr-Schmidt approach.[7]

However, comparing HSP Equilibria with the other approaches seems to be a little bit unfair because the existence of three types of actors leaves more room for "intermediate" probabilities of choosing *Left* or *Enter*. Therefore, we can carry out another test that does not have such a competitive advantage in favor of HSP Equilibria. This is done by counting the number of "correct predictions". By this we mean that one of the two following conditions holds: (a) Both the predicted probabilities of *Enter* (or *Left*) *and* the real percentage of *Enter* (or *Left*) are greater than or equal to 0.5; (b) Both the predicted probabilities of *Enter* (or *Left*) *and* the real percentage of *Enter* (or *Left*) are smaller than or equal to 0.5.

It shows that HSP Equilibria are "correct" in 46 out of 47 cases, Fehr-Schmidt predictions are correct in 39 cases and Charness-Rabin predictions are correct in 37 cases. Again, HSP Equilibria outperform the other concepts.

Let us now turn to the case where we take the worst HSP Equilibrium in each game. Of course, in this case R^2 and MAE decrease. However, $R^2 = 0.19$ and $MAE = 0.19$ are still better than the corresponding values of "best" Fehr-Schmidt and Charness-Rabin equilibria. Finally, if we only take into account games with unique HSP Equilibria then $R^2 = 0.698$ and $MAE = 0.13$ which are quite close to the case of the "best" equilibria.

In sum, the concept of HSP Equilibria is able to organize the data for behavior in the experiments of Charness and Rabin (2002) quite well and seems to be clearly superior to Nash Equilibria and the equilibria resulting from the approaches of Fehr-Schmidt and Charness-Rabin. It remains unclear, however, whether HSP Equilibria do as well in other games that have not been created for the special purpose of analyzing social preferences. So let us now look at some other games.

---

[7] I also tested the concept of Logit Equilibria (Anderson, Goeree and Holt 1997). In this case $R^2 = 0.23$ and $MAE = 0.203$. Furthermore, I combined the Logit Equilibria concept with the Charness-Rabin model. This lead to $R^2 = 0.2795$ and $MAE = 0.186$. Consequently, it seems that the idea of stochastic game theory helps increasing predictive accuracy. However, in both cases the HSP equilibrium still seems to be superior.

4. Treasures and Contradictions? An Application to the Goeree-Holt-Games

Recently, Jacob Goeree and Charles Holt (2001) analyzed ten pairs of games. Each pair consisted of two "similar" games with identical Nash Equilibria but differences in the absolute magnitude of payoffs. Goeree and Holt showed that in each case in one of the two versions of the game Nash Equilibria described laboratory behavior very well (these are called the "treasures") and that in the other version (the "contradictions") Nash Equilibria made very poor predictions. In this section we shall analyze eight of the ten pairs of games, again.[8] We use the same parameters of ρ_t, σ_t and θ_t and the same distribution of types as in section 3. It will be shown that HSP Equilibria can (partially) solve the puzzle that Goeree and Holt have found.

4.1. The One-Shot Traveler's Dilemma Game

In this section, let us consider the following game: Two players independently choose an integer number (N) between 180 and 300. If both players choose the same number they are paid this amount in money. Otherwise each player gets the minimum of both numbers plus (minus) a transfer payment (T) from the player who chose the higher number to the player with the lower number. Let $T > 1$.

The standard Nash Equilibrium of this game is that both players choose 180. Otherwise each has an incentive to underbid the other so that he can get the transfer payment. This Nash Equilibrium is unique and, furthermore, it is independent of T.

Goeree and Holt (2001, 1405-6) carried out two laboratory treatments, one with $T = 5$ and another with $T = 180$. It showed that in the latter case laboratory behavior was close to Nash Equilibrium. About 80 percent chose numbers very close to 180. In contrast to this, laboratory subjects did not at all behave according to the Nash prediction in the treatment with $T = 5$. Here, even slightly more than 80 percent of laboratory subjects chose numbers that were close to the maximum, 300, which is not part of any Nash Equilibrium.

Next, consider how HSP Equilibria correspond to laboratory behavior. Let us begin with the $T = 180$ treatment. It shows that in this case HSP works as well as Nash Equilibria because both coincide:

---

[8] The remaining two pairs of games that deal with incomplete information games are not too interesting here because the concept of HSP Equilibria always has the character of incomplete information games anyway.

Result 1: *The one shot traveler's dilemma game with $T = 180$ has a unique HSP Equilibrium in which all types, SE, IA and WP actors, choose 180 regardless of their roles as player 1 or player 2.*

Sketch of the Proof:
(a) One can easily verify that no player and no type has an incentive to deviate from the Equilibrium. Therefore it constitutes an equilibrium.
(b) Assume that there exists another HSP Equilibrium in which $N > 180$ is played with positive probability. Let N_{max} be the highest number of the equilibrium candidate that is played with positive probability. First, we can easily rule out that it could be an equilibrium that all types play the same number $N > 180$ with probability one as it is always advantageous for SE players to underbid the others by one unit. Second, it can be shown that regardless of the strategies of IA and WP players underbidding incentives of SE players are so strong that only $N = 180$ remains an equilibrium candidate for SE actors. Third, given that SE players always play 180 and that their share of the population is 0.5, it can be shown that no IA player would choose to play N_{max} with positive probability because they always prefer playing 180 to N_{max}. Furthermore, it can be shown that expected utility of IA subjects playing $N > 180$ is always negative whereas they can realize a strictly positive expected utility if they choose 180. Consequently, $N = 180$ remains the only candidate for equilibrium behavior of IA players. Fourth, given that SE and IA players play 180, the best remaining scenario for WP subjects playing $N > 180$ gives them an expected utility of 175.2 which is strictly less than 180, the expected utility of playing $N = 180$. Thus, there cannot exist another equilibrium in which $N > 180$ is played with strictly positive probability. □

Let us now turn to the case $T = 5$. Here we have multiple HSP Equilibria:

Result 2: *In the one shot traveller's dilemma game with $T = 5$ there exists more than one HSP Equilibrium.*
(1) One Equilibrium is that all types choose $N = 180$.
(2) Another HSP Equilibrium contains mixed strategies. Here, WP players choose $N_{WP} = 300$ and IA players choose $N_{IA} = 288$. In both cases players choose their numbers with probability one. In contrast to this, SE subjects play a mixed strategy with (rounded) probabilities:
$P(299) = 0.06$; $P(298) = 0.092$; $P(297) = 0.0784$; $P(296) = 0.1077$; $P(295) = 0.1$; $P(294) = 0.1276$; $P(293) = 0.1255$; $P(292) = 0.1527$; $P(291) = 0.1560$.

The second equilibrium corresponds well to the behavior of subjects in the laboratory. The intuition is that WP and IA players play some kind of a coordination game. With $T = 5$ it is

sufficient for WP players that the other WP players choose 300, regardless of how the other types behave. In this case they prefer $N_{WP} = 300$. IA subjects have a very strong aversion of getting less than other players. In general they could coordinate their behavior on any number that is smaller than (or equal to) SE and WP players' choices. However, only $N_{IA} = 288$ makes SE subjects that play their mixed strategies (given above) indifferent to all numbers between 291 and 299. Finally, SE players experience a tradeoff between efficiently underbidding WP players (299) and between being underbidden by IA subjects. However, because 288 is sufficiently below $N_{WP} = 300$ and because WP players have a much larger share in the population it pays for SE subjects to risk being exploited by IA subjects. Note that this equilibrium works only if $T = 5$, i.e. it is important for this equilibrium that being underbidden by others is not too costly. This is why WP subjects can coordinate on 300 and SE players risk being underbidden by IA subjects.

Summarizing, HSP Equilibria explain laboratory behavior in the traveller's dilemma in both treatments with low and high transfer payments quite well.

4.2. Matching Pennies Games

Three variations of a "Matching Pennies" game are the subject of this section. Table 4 gives the basic structure of the games in which the players have to move simultaneously.

		Player 2	
		Left	*Right*
Player 1	*Top*	A,40	40,80
	Bottom	40,80	80,40

Table 4: Matching Pennies Games

In the symmetric game $A = 80$, in the asymmetric game $A = 320$ and in the reverse asymmetric game $A = 44$. In all three variations there does not exist a Nash Equilibrium in pure strategies. In particular, in the symmetric game the equilibrium in mixed strategies consists of both subjects playing each strategy with equal probability. Note that this behavior remains the Nash Equilibrium strategy for player 1 in the other variations, too.

Nash Equilibrium describes laboratory behavior in the symmetric case very well. Here, player 1 plays *Top* in 48 percent of all subjects and player 2 chooses *Left* in 48 percent, too. In contrast to this, Nash Equilibrium fails to explain player 1's behavior in the other cases. In the asymmetric game 96 percent of subjects chose *Top* and 84 percent of players 2 selected *Right*. In the reversed asymmetric game 92 percent of the row players took Bottom and 80 percent of

the column players decided to take *Left*. Nash Equilibrium thus only explains behavior in the symmetric game which leads Goeree and Holt (2001, 1407) to summarize: "… the Nash mixed-strategy prediction seems to work only by coincidence ..."

Let us now turn to HSP Equilibria. In the symmetric game we get the following equilibria:

Result 3: *In the symmetric "Matching Pennies Game" all strategy combinations are HSP Equilibria which fulfill the following conditions:*

$0.5 \cdot p_{SE}[Left] + 0.15 \cdot p_{IA}[Left] + 0.35 \cdot p[Left] = 0.5$ *and*

$0.5 \cdot p_{SE}[Top] + 0.15 \cdot p_{IA}[Top] + 0.35 \cdot p[Top] = 0.5$.

For example, if all types mix *Left-Right* or *Top-Bottom* with probabilities (0.5, 0.5) this constitutes a HSP Equilibrium. The same is true for SE subjects choosing *Top* or *Left* and the other types choosing *Bottom* of *Right*. In any case, the aggregate probability of choosing *Left* or *Top* must be 0.5. Obviously, this explains behavior exactly as good as Nash Equilibrium does.

Consider now the asymmetric game. In this case column players of type WP have a dominant strategy (*Left*) so that one can derive the next HSP Equilibrium that differs fundamentally from the Nash Equilibrium.

Result 4: *In the asymmetric "Matching Pennies Game" there exists a unique HSP Equilibrium with* $(p_{SE}[Top], p_{IA}[Top], p_{WP}[Top], p_{SE}[Left], p_{IA}[Left], p_{WP}[Left]) = (1,1,1,0,0,1)$,

i.e. all row players choose Top, SE and IA column players choose Right and WP column players choose Left.

This means that the equilibrium aggregate probability of playing *Top* equals 1 (experimental behavior: 96 percent) and the aggregate probability of playing *Left* is 0.35 (experimental behavior: 16 percent). In contrast to Nash Equilibrium the concept of HSP Equilibria reacts to the payment variation!

A similar picture can be drawn in the reversed asymmetric game:

Result 5: *In the reversed asymmetric "Matching Pennies Game" there exists a unique HSP Equilibrium with*

$$(p_{SE}[Top], p_{IA}[Top], p_{WP}[Top], p_{SE}[Left], p_{IA}[Left], p_{WP}[Left]) = \left(\frac{199}{670},1,0,1,1,\frac{57}{77}\right).$$

Consequently, the aggregate probability that row players play *Top* equals 0.30 (compared with 8 percent in the laboratory) and the aggregate probability of *Left* is 0.91 (experimental

probability: 80 percent). Again, HSP Equilibria explain the main qualitative features of behavioral change due to the variations of payoffs. Once more, HSP Equilibrium turns a "contradiction" into a "treasure".

4.3. A Coordination Game with a secure outside option

In this section we analyze coordination games with a secure outside option. However, the outside option is dominated by a mixed strategy of *Left* and *Right* so that it should never be part of a Nash Equilibrium. Table 5 gives the structure of the games:

		Player 2		
		Left	*Right*	*Secure*
Player 1	*Top*	90,90	0,0	*A*,40
	Bottom	0,0	180,180	0,40

Table 5: Coordination Games with outside option

There are two versions of the game. In the first $A = 0$ and in the second $A = 400$. The games have three equilibria: (1) (*Bottom*,*Right*), (2) (*Top*,*Left*) and an equilibrium in mixed strategies (3) $\left(\left(p_{Left}, p_{Right}\right), p_{Top}\right) = \left(\left(\frac{2}{3}, \frac{1}{3}\right), \frac{2}{3}\right)$. Note that the Nash Equilibria are independent of A. In the $A = 0$ treatment 96 percent (84 percent) of the row players (column players) have chosen *Bottom* (*Right*), i.e. the strategies for the pareto dominant Nash Equilibrium. 80 percent of the subjects managed to coordinate on this equilibrium. This was different in the $A = 400$ treatment. Here only 64 percent (76 percent) of the row (column) players have chosen *Bottom* (*Right*) and only 32 percent of the pairs coordinated on this equilibrium. More than 50 percent of the outcomes were uncoordinated non-Nash outcomes. Again, Nash Equilibrium was a bad predictor for one of the versions ($A = 400$).

It turns out that HSP Equilibria in the $A = 0$ treatment coincide with Nash Equilibria, i.e. in the "treasure" version HSP Equilibria do equally well. However, HSP Equilibria in the $A = 400$ version differ from Nash Equilibria.

Result 6: *There are three HSP Equilibria in the extended coordination game with A = 400:*
(1) SE, IA and WP players play (Bottom, Right).
(2) All types of row players play Top. SE and IA column players choose Left and WP column players choose Secure.
(3) Row players: SE and IA subjects play Top and WP subjects play Top with probability

1/21. Column players: SE and IA subjects play Left with probability 442/1755 and Right with probability 1313/1755. WP subjects choose Secure.

The third HSP Equilibrium gives an aggregate probability that players choose *Top* of 2/3. The aggregate probability of *Left* is about 0.16, the probability of *Right* (*Secure*) is 0.49 (0.35). Consequently, only about 27 percent of the pairs can be expected to coordinate on one of the two Nash Equilibria. Obviously, this estimate looks more pessimistic than the experimental experience. Nevertheless it better fits the data than Nash Equilibria.

4.4. A Minimum-Effort Coordination Game

This game is a special kind of a team problem. Each of the two players simultaneously chooses his effort e and payoffs are determined by the formula $y_i = \min\{e_i, e_j\} - ce_i$, $i, j = 1, 2$ and $i \neq j$. Here c represents a cost parameter that is assumed to be smaller than 1 and ce_i gives individual i's costs. Each individual gets the output $\min\{e_i, e_j\}$. It is straightforward that there are multiple Nash Equilibria. In fact, every feasible effort is part of an equilibrium if all actors coordinate on the corresponding value. This is true for any $c < 1$ and is independent from the magnitude of c.

In the experimental design by Goeree and Holt (2001) efforts could be any integer number between 110 and 170. They carried out two treatments, one with $c = 0.1$ and one with $c = 0.9$. It shows that behavior in these two treatments clearly differs. With $c = 0.1$ choices of effort concentrate near the upper boundary, 170. In contrast to this, most subjects have chosen efforts near 110, the lower boundary, if $c = 0.9$. Without doubt, both treatments are in accordance with Nash Equilibrium. However, the concept of Nash Equilibrium gives no hints why there is so much divergence between the treatments.

Unfortunately, there are also multiple HSP Equilibria. In fact, if all types coordinate on *any* feasible effort this represents a HSP Equilibrium, too. Consequently, as with Nash Equilibrium, HSP Equilibria are in accordance with experimental behavior. This, of course, is not surprising because any effort choice is part of one of the many Nash Equilibria of the game. Even worse, the behavioral differences between the two treatments cannot be explained, either. One might argue, however, that with $c = 0.1$ the coordination problem is weakened. For example, even if all SE and IA players choose $e = 110$ it is an optimal behavior for WP players to coordinate on $e = 170$, i.e. it is sufficient for WP subjects that coordination only between them is arranged properly. Given that all WP players coordinate on $e = 170$ it is optimal for all SE players to coordinate on 170, too, even if all IA subjects stick to $e = 110$. Finally, as

all types playing 170 is a HSP Equilibrium, IA players would follow the other types of players. Nevertheless, such a kind of reasoning is not part of the concept of HSP Equilibrium so that we have an unresolved equilibrium selection problem.

4.5. The Kreps Game

In the previous section we dealt with a game that had multiple equilibria. Every feasible effort could be part of a Nash Equilibrium and there was no reliable way to discriminate between the Nash Equilibria. In this section Nash even works worse. There are, again, multiple Nash Equilibria but for one player the only pure strategy which is not part of any Nash Equilibrium is chosen most of the times in one of the treatments. The structure of the Kreps Game is given in Table 6.

	Left	*Middle*	*Non-Nash*	*Right*
Top	200,50	0,45	10,30	20,–250
Bottom	0,–250	10,–100	30,30	I: 50,40 (II: 350,400)

Table 6: Two variants of the Kreps Game

The only difference between the variants can be found in the cell (*Bottom,Right*). In variant I payoffs are rather small and in variant II they are much larger. However, both variants of the Kreps Game have two Nash Equilibria in pure strategies, (*Top,Left*) and (*Bottom,Right*). Furthermore there exists one Nash Equilibrium in mixed strategies in which the column player randomizes between *Left* (with probability 1/21) and *Middle* (with probability 20/21) and the row player randomizes between *Top* (with probability 150/155) and *Bottom* (with probability 5/155). Consequently, only the pure column strategy *Non-Nash* is not part of a Nash Equilibrium in both variants of the game.

Here, variant II represents the treasure treatment in which 96 percent of the subjects in the laboratory have chosen *Bottom* and 84 percent have chosen *Right*. So behavior is in accordance with the pareto optimal Nash Equilibrium. However, in treatment I with relatively low payoffs behavior of subjects in the laboratory was completely at odds with the concept of Nash Equilibrium. 68 percent of the column players have chosen *Non-Nash*, the only strategy that is not part of any Nash Equilibrium.

In contrast to Nash Equilibrium HSP Equilibrium does not rule out the pure strategy *Non-Nash*:

Result 7: *Variant I of the Kreps Game has multiple HSP Equilibria. One of them is characterized by the following behavior: SE and IA column players choose Non-Nash and WP column players choose Non-Nash with probability 57/77 and Left with probability 20/77. SE row players choose Top with probability 107/140 and Bottom with probability 33/140. IA row players choose Bottom and WP row players choose Top.*

According to this HSP Equilibrium, *Non-Nash* is played with an aggregated probability of approximately 91 percent (compared to 68 percent in the laboratory) and the probability of *Left* is about 9 percent (26 percent). The aggregate probability of choosing *Top* is 73.2 percent (compared to 68 percent in the laboratory). HSP Equilibrium thus gives a clearly better prediction than Nash Equilibrium. It should be mentioned, however, that the HSP Equilibrium from Result 7 is also valid for treatment II where it is hardly played at all. In addition, note that only in variant II there exists a HSP Equilibrium in which *all types* of row players choose *Bottom* and *all types* of column players choose *Right*.

4.6. Should you trust others to be rational?

Let us now turn to games in the extensive form. What is of particular interest here is whether the logic of backward induction holds reliably. This means that we have to analyze whether players who move first should trust their followers to behave rational and whether they should believe threats that are not credible. In this section we concentrate on the first question. Look at the game in Figure 3 in which the first player has to decide whether to stop the game and choose a safe payoff or whether he should let the second player choose between two other payoff combinations.

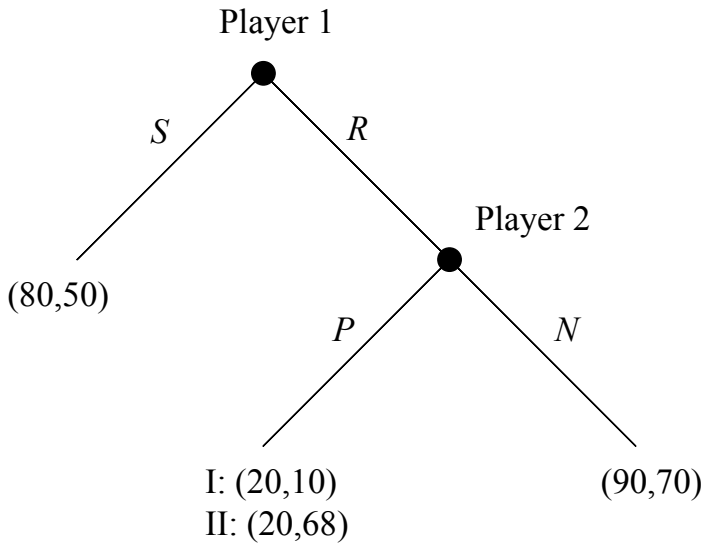


Figure 3: An extensive form game

Again, there are two variants of the game. The difference between the variants consists of the payoff of player 2 in case of an *R-P* play. In variant I player 2 looses much money if he chooses *P* instead of *N*. In variant II payoff differences for player 2 are rather small if he has to choose between *P* and *N*. In any case there is the same unique subgame perfect Nash Equilibrium, namely (*R,N*).

Goeree and Holt (2001) show that laboratory behavior fits well to the Nash prediction in variant I: 84 percent of the first movers have chosen *R* and 100 percent of the second movers selected *N*. However, things are quite different in variant II. Here only 48 percent of the first movers decided to take *R* and 75 percent of the second movers have chosen *N*. Although most second movers behaved rationally, there are sufficiently many subjects who deviate from the rational second move so that it paid on average for the first movers to choose the non-equilibrium strategy *S*. Or to put it another way: In variant II the first movers have good reason *not* to trust the other players to behave rationally. Again, Nash Equilibrium does well in one treatment but fails in the other one.

In variant I, the treasure treatment, the HSP Equilibrium coincides with the Nash Equilibrium. Consequently, HSP Equilibria are exactly as successful in this variant as Nash Equilibria. However, variant II has a different HSP Equilibrium.

Result 8: *HSP Equilibrium play in variant II of the extensive form game is characterized by the following behavior: SE and IA types of the first movers choose S and WP subjects take R. The second movers choose N if they are SE or WP types. They prefer P if they are IA actors.*

The intuition for this result is that inequity averse players prefer *P* to *N*. The reason for this is that they dislike being in the disadvantaged position much more than they dislike inequity in their own "favor".[9] Because IA second movers deviate from the subgame perfect Nash behavior SE and IA subjects prefer *S* as first movers.

Consequently, HSP Equilibrium predicts that 65 percent of the first movers choose *S* (compared to 52 percent in experimental behavior) and only 35 percent choose *R*. Furthermore, according to HSP Equilibrium 15 percent of the second movers opt for *P* (compared to 25 percent in the experiments). Again, HSP Equilibrium is a better estimator of actual behavior than subgame perfect Nash equilibrium.

---

[9] This, of course, is due to the parameters for inequity aversion taken from Fehr, Klein and Schmidt (2001) and Fehr, Kremhelmer and Schmidt (2002).

Finally, note that the section title "should you trust others to be rational" that has been adopted from Goeree and Holt (2001) does not really fit to the game when looked upon from a HSP perspective. In equilibrium all players behave rationally, so that one can trust in the others rationality. However, what the first movers do not know is the motivation of the second movers. Consequently, from a HSP point of view the problem considered here is better described by the question "should you 'trust' others not to be envious?"

4.7. Should you believe a threat that is not credible?

In this section we deal with the problem of credible vs. incredible threats. Many Nash Equilibria involve incredible threats that make them rather implausible. For this reason Selten (1965) introduced the criterion of subgame perfectness to rule them out. However, it has often been shown that subgame perfectness does not always fit to actual behavior very well. Let us now look at two games with incredible threats that are represented in Figure 4.

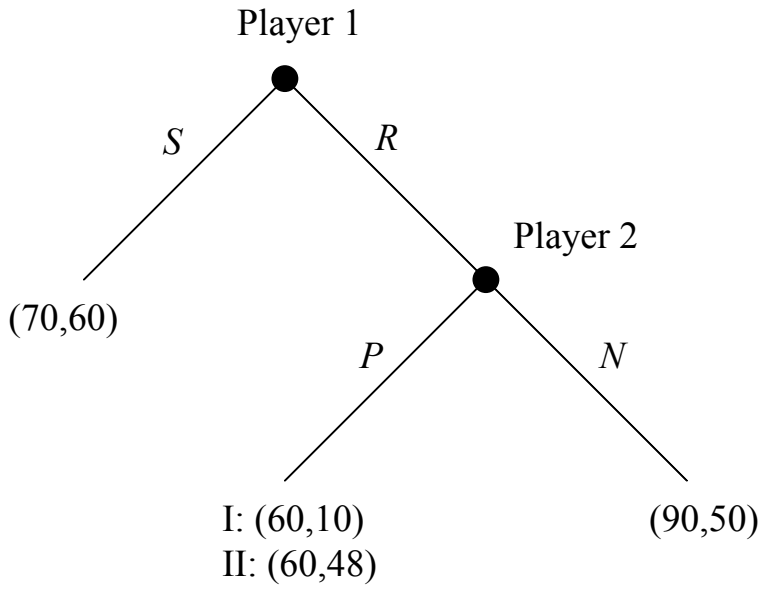


Figure 4: Two extensive form games with incredible threats

The structure of the games are very similar to those in the previous section. However, in the games considered here player 2 dislikes player 1 to play *R*. The Nash Equilibrium (*S,P*) works only with the use of the incredible threat of player 2 to play *P*. Subgame perfectness rules this out so that (*R,N*) remains the unique subgame perfect Nash Equilibrium.

Again, the only difference between the two variants of the game consists of different payoffs in case of (*R,P*) play. In variant I playing *P* is very costly to player 2 because he looses 80 percent of his profits. In contrast to this variant II is a game in which playing *P* only costs 4

percent of maximum payoffs for player 2 although this has absolutely no impact on the subgame perfect Nash Equilibrium of the game.

Goeree and Holt (2001) find that variant I is a treasure, again. In their experiments 88 percent of the first movers have chosen *R* and 100 percent of the second movers have chosen *N*. This nearly perfectly fits to the prediction of the subgame perfect Nash Equilibrium. Things look very different in variant II of the game. Here only 68 percent of the first movers have chosen *R* and only 53 percent of the second movers have chosen *N*, the subgame perfect equilibrium strategy.

In variant I (the treasure treatment) of the game HSP Equilibria are even more precise than subgame perfect Nash Equilibria.

Result 9: *In variant I of the extensive form game with incredible threats there exists a unique HSP Equilibrium. Here all types of second movers choose N. First movers that are SE and WP types choose R and IA first movers choose S.*

Consequently, the aggregate probability of first movers choosing *R* is 85 percent (compared to 88 percent in laboratory experiments) and the percentage of second movers playing *N* is 100 percent (compared to 100 percent in the experiments). In variant II of the game we get a different HSP Equilibrium.

Result 10: *The unique HSP Equilibrium of variant II of the extensive form game with incredible threats is given by the following behavior: SE- and WP-first movers choose R and IA-first movers choose S. If second movers are of SE or WP type they choose N and if they are IA-subjects they choose P.*

This means that 85 percent (compared to 68 percent in the experiment) of all first mover decisions should be *R*, again, and that 85 percent (compared to 53 percent) of the second movers are expected to choose *N*. Obviously, this prediction is worse than the one from variant I. It seems that HSP Equilibrium does not react sufficiently strong to the change in parameter values. However, it still is clearly superior to the prediction according to the subgame perfect Nash Equilibrium.

Finally, from a HSP point of view the section title which has again been adopted from Goeree and Holt (2001) poses the wrong question. It is not whether you should believe an incredible threat or not, it is much more the question *whether* a threat is credible or not, given that there are heterogeneous actors.

4.8. Two-Stage Bargaining Games

Bargaining games have always been a very special challenge to economic theory in general and game theory in particular. In this section we analyze a two-stage alternating offers bargaining game. In stage 1 player 1 makes his first proposal (x_1) about how to split a pie of \$5 and player 2 then decides whether to accept this proposal or not. If the offer is accepted both players get the proposed amount of money. However, if it is rejected the pie shrinks to \$2 (variant I) or \$0.5 (variant II). In this case player 2 gets the right to propose the split of the pie (x_2) and player 1 can accept or reject this second offer. If the second offer is also rejected both players get a payoff of zero. If player 1 agrees both players get the proposed amounts of money.

This game has the following subgame perfect equilibrium: In stage 2 player 2 proposes that he gets all the remaining money but 1 cent. Player 1 accepts this proposal. Anticipating this, player 1 offers player 2 an amount of money that is equal to the magnitude of the pie in stage 2 and keeps the rest for himself. This proposal is also accepted by player 2. Consequently, in variant I player 1 gets an equilibrium payoff of \$3 and player 2 gets \$2. In contrast to this, in variant II equilibrium play is characterized by a distribution in which player 1 gets \$4.50 and player 2 gets \$0.50.

Again, it turns out that experimental behavior in one of the treatments is close to the subgame perfect Nash Equilibrium and in the other the equilibrium is a very poor predictor of behavior. In variant I, the treasure treatment, the average demand of player 1 is \$2.83 which is quite close to the proposed \$3. In variant II, however, average demand of player 1 in stage 1 is only \$3.38 which is far below the game theoretical prediction of \$4.50.

In both variants of the game HSP Equilibria do not coincide with the subgame perfect Nash Equilibrium. In variant I it is quite close to Nash Equilibrium play, though.

Result 11: *HSP Equilibria of variant I of the two stage bargaining game are characterized by the following behavior of player 1: SE types demand \$3 and accept any offer in stage 2. IA subjects demand \$2.5 and accept player 2's offer in stage 2 if* $x_2 \leq 6/5$. *WP players demand \$2.50 and accept any offer in stage 2. HSP Equilibrium strategies of player 2 are given by: SE players accept the offer in stage 1 if* $x_1 \leq 3$ *and demand* $x_2 = 2$ *in stage 2. IA subjects only accept the single offer* $x_1 = 2.5$ *and demand* $x_2 = 2$ *in stage 2. WP players accept all offers in stage 1 and demand* $1 \leq x_2 \leq 2$ *in stage 2.*

As a consequence, the average demand of player 1 in stage 1 is expected to be \$2.75. This is even closer to the experimental behavior (\$2.83) than the subgame perfect Nash Equilibrium prediction. The intuition is that IA subjects prefer an equal split. WP subjects also choose this particular demand because this makes sure that the offer is not rejected by any type. In equilibrium play all IA and WP offers are accepted. SE players risk demanding \$4.5 because they know that only IA actors reject this offer. In addition, IA players who have rejected an offer demand the whole pie in stage 2 because of negative reciprocity. This offer, however, is accepted by SE and WP players. IA players would reject it. However, this never happens because IA types of player 1 make offers that are always accepted in stage 1.

Result 12: H*SP Equilibria of variant II of the two stage bargaining game are characterized by the following behavior of player 1: SE types demand \$4.50 and accept any offer in stage 2. IA subjects demand \$2.50 and accept player 2's offer in stage 2 if* $x_2 \leq 3/10$. *WP players demand any amount between \$2.50 and \$2.70 and accept any offer in stage 2. HSP Equilibrium strategies of player 2 are: SE players accept the offer in stage 1 if* $x_1 \leq 4.5$ *and demand* $x_2 = 0.5$ *in stage 2. IA subjects only accept offers that are below* $x_1 = 2.70$ *and demand* $x_2 = 0.50$ *in stage 2. WP players accept all offers in stage 1 and demand* $0.25 \leq x_2 \leq 0.50$ *in stage 2.*

Because of the multiplicity of equilibria given in Result 12 we do not have a unique expected value of the first demand in stage 1. The mean demand should be between \$3.5 and \$3.57 which is quite close to the behavior in the laboratory (\$3.38). As we have mentioned before the subgame perfect Nash Equilibrium gives a clearly worse prediction of first proposals. The difference between the equilibria of variant I and II stems from the more aggressive play of SE actors. However, WP players only change their strategies slightly and IA types do not change their first offers at all. The latter point explains why experimental behavior does not change as much as standard game theory lets us expect.

5. Conclusion

The main purposes of this paper are to demonstrate the usefulness of social preferences in explaining economic behavior and, even more important, to show that *heterogeneity* of preferences can play an *important* role in explaining many deviations of laboratory behavior from standard game theoretical predictions.

For these purposes, we introduce the concept of HSP Equilibrium and show that – compared with other well known approaches – it is able to explain behavior of subjects in the experi-

ments of Charness and Rabin (2002) very well. The idea of HSP Equilibria integrates the competing approaches of Fehr and Schmidt (1999) and Charness and Rabin (2002) into a unified and tractable framework. According to HSP Equilibrium three types of players, strictly egoistic subjects, inequity averse agents and subjects with (social) welfare preferences, behave according to the corresponding (perfect) Bayesian Equilibria.

HSP Equilibrium predictions are clearly superior to Nash Equilibrium, the inequity aversion model (Fehr and Schmidt 1999) and Charness' and Rabin's model of Social Welfare Preferences. Furthermore, it was shown that HSP Equilibrium can explain most "behavioral anomalies", the "contradictions", that have been presented in Goeree and Holt (2001). The overall impression is that the analytical combination of social preferences with heterogeneity in these preferences is a very fruitful approach to understand real behavior that is observed in laboratory experiments.

HSP Equilibrium also explicitly takes into account negative reciprocity. Although this helps to get better predictions in a few games reciprocity does not play a major role in most games that have been considered here. Therefore, it remains unclear whether this achievement justifies the resulting analytical inconvenience.

Finally, the HSP Equilibrium approach is far from being perfect, of course. There still remains an uncomfortably high level of "unexplained variation". So where do we go from here? Presumably, there are two different ways one may try to make progress. The first one is to allow for even more than the three different types that are used in the HSP Equilibrium concept. For example, one can add subjects with competitive preferences, i.e. actors who always put negative weights on other people's payoffs. The author has tried this procedure. However, it showed that the introduction of preferences did not improve the predictive success of the approach. Furthermore, the more types are integrated the more tedious, i.e. less applicable, the analysis becomes. Therefore, the author is quite pessimistic about this alternative for future research.

Another way to proceed is to try to integrate heterogeneous social preferences in the Quantal Response Equilibrium framework. In an intuitive appealing manner the proponents of this approach, in particular McKelvey and Palfrey as well as Goeree and Holt and their coauthors, show that they can explain the behavioral consequences of parameter variations in quite a lot of games and experiments. Nevertheless, there is still much room for improvements in their framework, too. Consequently, it looks like a natural next step to combine the idea of noisy decision making with the approach of heterogeneous social preferences.

References

Anderson, S.P., Goeree, J. and Holt, Ch.A. (2002), The Logit Equilibrium: A Perspective on Intuitive Behavioral Anomalies, *Southern Economic Journal*, 69(1), pp. 21-47.

Andreoni, J. and Miller, J. (2002), Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism, Econometrica, 70, 737-753.

Bolton, G. and Ockenfels, A. (2000), A Theory of Equity, Reciprocity and Competition, American Economic Review, 100, S. 166-193.

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

Dufwenberg, M. and Kirchsteiger, G. (1998), A Theory of Sequential Reciprocity, Working Paper, CentER, University of Tilburg.

Fehr, E. and Schmidt, K. (1999), A Theory of Fairness, Competition and Co-operation, Quarterly Journal of Economics, 114, S. 817-868.

Fehr, E. and Schmidt, K.M. (2000), Fairness, Incentives, and Contractual Choices, European Economic Review, 44, S. 1057-1068.

Fehr, E., Klein, A. and Schmidt, K.M. (2001), Fairness, Incentives and Contractual Incompleteness, Working Paper, University of Munich.

Fehr, E., Kremhelmer, S. and Schmidt, K.M. (2002), Fairness and the Optimal Allocation of Ownership Rights, Working Paper, University of Munich.

Fudenberg, D. and Levine, D.K. (1998), The Theory of Learning in Games, MIT Press: Cambridge and London.

Goeree, J. and Holt, Ch. (2001), Ten Little Treasures of Game Theory and Ten Intuitive Contradictions, American Economic Review, 91, S. 1402-1422.

Kagel, J. and Wolfe, K. (2000), Tests of Difference Aversion to Explain Anomalies in Simple Bargaining Games, mimeo, Ohio State University.

McKelvey, R.D. and Palfrey, T.R. (1995), Quantile Response Equilibria for Normal Form Games, Games and Economic Behavior, 10, 6-38.

McKelvey, R.D. and Palfrey, T.R. (1998), Quantal Response Equilibria in Extensive Form Games, Experimental Economics, 1(1), 9-41.

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

Selten, R. (1965), Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit, Zeitschrift für die gesamte Staatswissenschaft, 12, 301-324.

Weibull, J. (1995), Evolutionary Game Theory, MIT Press: Cambridge and London.